Water Resources Research

RESEARCH ARTICLE
10.1029/2019WR026085

Special Section:
Advancing process representation in hydrologic models: Integrating new concepts, knowledge, and data

Improving the Predictive Skill of a Distributed Hydrological Model by Calibration on Spatial Patterns With Multiple Satellite Data Sets

Moctar Dembélé[1,2], Markus Hrachowitz[2], Hubert H. G. Savenije[2], Grégoire Mariéthoz[1], and Bettina Schaefli[1,3]

[1]Faculty of Geosciences and Environment, Institute of Earth Surface Dynamics, University of Lausanne, Lausanne, Switzerland, [2]Faculty of Civil Engineering and Geosciences, Water Resources Section, Delft University of Technology, Delft, Netherlands, [3]Now at Institute of Geography (GIUB), University of Bern, Bern, Switzerland

Key Points:
- A spatial calibration framework is developed by combining four noncommensurable variables describing different hydrological processes
- A new bias-insensitive metric is developed to incorporate spatial heterogeneity in a multivariate calibration scheme
- Only spatial patterns of satellite data are used to improve the predictive skill of a distributed hydrological model

Supporting Information:
- Supporting Information S1

Correspondence to:
M. Dembélé,
moctar.dembele@unil.ch;
m.dembele@tudelft.nl

Citation:
Dembélé, M., Hrachowitz, M., Savenije, H. H. G., Mariéthoz, G., & Schaefli, B. (2020). Improving the predictive skill of a distributed hydrological model by calibration on spatial patterns with multiple satellite data sets. *Water Resources Research*, *56*, e2019WR026085. https://doi.org/10.1029/2019WR026085

Received 1 AUG 2019
Accepted 13 JAN 2020
Accepted article online 15 JAN 2020

Abstract Hydrological model calibration combining Earth observations and in situ measurements is a promising solution to overcome the limitations of the traditional streamflow-only calibration. However, combining multiple data sources in model calibration requires a meaningful integration of the data sets, which should harness their most reliable contents to avoid accumulation of their uncertainties and mislead the parameter estimation procedure. This study analyzes the improvement of model parameter selection by using only the spatial patterns of satellite remote sensing data, thereby ignoring their absolute values. Although satellite products are characterized by uncertainties, their most reliable key feature is the representation of spatial patterns, which is a unique and relevant source of information for distributed hydrological models. We propose a novel multivariate calibration framework exploiting spatial patterns and simultaneously incorporating streamflow and three satellite products (i.e., Global Land Evaporation Amsterdam Model [GLEAM] evaporation, European Space Agency Climate Change Initiative [ESA CCI] soil moisture, and Gravity Recovery and Climate Experiment [GRACE] terrestrial water storage). The Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature data set is used for model evaluation. A bias-insensitive and multicomponent spatial pattern matching metric is developed to formulate a multiobjective function. The proposed multivariate calibration framework is tested with the mesoscale Hydrologic Model (mHM) and applied to the poorly gauged Volta River basin located in a predominantly semiarid climate in West Africa. Results of the multivariate calibration show that the decrease in performance for streamflow (−7%) and terrestrial water storage (−6%) is counterbalanced with an increase in performance for soil moisture (+105%) and evaporation (+26%). These results demonstrate that there are benefits in using satellite data sets, when suitably integrated in a robust model parametrization scheme.

1. Introduction

One of the key challenges in hydrological modeling (Beven, 2019a; Singh, 2018) is the reliable representation of the spatiotemporal variability of natural processes, to which the footprint of human activity is often superimposed. In most places, available in situ observations are not sufficient to capture the spatiotemporal heterogeneity of dominant hydrological processes (AghaKouchak et al., 2015; Hrachowitz & Clark, 2017). With the upswing in development of distributed hydrological models (DHMs) that offer spatially explicit predictions as an essential tool for decision making (Fatichi et al., 2016; Kampf & Burges, 2007; Paniconi & Putti, 2015; Semenova & Beven, 2015), there is a growing interest in the plausibility of their spatial patterns (Ko et al., 2019; Koch et al., 2018; Stisen et al., 2018; Wealands et al., 2005; Zink et al., 2018).

Most commonly, hydrological models are calibrated using streamflow data alone (Becker et al., 2019; Yassin et al., 2017). The streamflow signal represents an integrated response of the hydrological system to a set of natural drivers (e.g., climate and landscape) and anthropogenic influences (e.g., deforestation and reservoirs) occurring upstream of the measurement's location (Koch et al., 2015; Rientjes et al., 2013). Although streamflow is key to understanding the temporal dynamics of a system, it does not disclose much information on the system-internal spatial heterogeneity of the hydrological processes (McDonnell et al., 2007; Rajib et al., 2018). It therefore has little discriminatory power to constrain the feasible parameter space

of a distributed model, i.e., the boundary flux or closure problem (Beven, 2006b). Consequently, a spatially DHM calibrated only on streamflow is very unlikely able to reproduce a reliable spatiotemporal representation of other hydrological fluxes and states (Birhanu et al., 2019; Clark et al., 2016; Grayson & Bloschl, 2001; Hrachowitz et al., 2014; Livneh & Lettenmaier, 2012; Minville et al., 2014), even if a multiscale parameter regionalization (MPR) scheme is used (Rakovec et al., 2016). Mismatches between temporal and spatial patterns should therefore be expected when comparing hydrological model outputs to other distributed observational data sets (Vereecken et al., 2008; Xu et al., 2014).

For a few decades, satellite remote sensing (SRS) has opened up new avenues for the development of spatial hydrology (Cui et al., 2018; Engman & Gurney, 1991; Lettenmaier et al., 2015; McCabe et al., 2017; Mendoza et al., 2002; Pasetto et al., 2018; Schmugge et al., 2002). The increasing and unprecedented availability of SRS data at increasingly finer spatial and temporal resolutions has triggered the development of large-domain water management applications including flood and drought monitoring (Hapuarachchi et al., 2011; Klemas, 2014; Revilla-Romero et al., 2015; Senay et al., 2015; Sheffield et al., 2012; Su et al., 2017; Teng et al., 2017; Wu et al., 2014). The use of SRS data in water resources monitoring is promising, and it has led to an increasing number of studies on a variety of topics in hydrology, including precipitation, evaporation, and soil moisture estimation (Cazenave et al., 2016; Chen & Wang, 2018; Cui et al., 2019; National Academies of Sciences, Engineering, and Medicine, 2019; Schultz & Engman, 2012). SRS data complement in situ hydrometeorological data (Balsamo et al., 2018), which are typically scarce and whose unavailability hinders the understanding of environmental systems (Tang et al., 2009). This aspect is particularly relevant for developing countries where research for development initiatives have been increasing in the recent years (Montanari et al., 2015).

Besides direct use of SRS data for water resources monitoring and management (Cui et al., 2019; Sheffield et al., 2018), an increasing body of literature addresses the question of how these data sets can be used to improve hydrological modeling (Baroni et al., 2019; Clark et al., 2015; Guntner, 2008; Liu et al., 2012; Nijzink et al., 2018; Paniconi & Putti, 2015; Parajka et al., 2009). The scientific community has, in fact, long been advocating the use of spatial data for DHM evaluation (Beven & Feyen, 2002; Grayson & Bloschl, 2001; Koch et al., 2015; Refsgaard, 2001; Wealands et al., 2005). SRS data sets have the potential to improve models either via data assimilation (Leroux et al., 2016; Tangdamrongsub et al., 2017; Tian et al., 2017) or via calibration (Bai et al., 2018; Li et al., 2018; Rientjes et al., 2013). In this context, data assimilation is used to update the states of a given model, e.g., to compensate for model structural deficiencies (Spaaks & Bouten, 2013). For parameter estimation (i.e., model calibration) with SRS data, the existing approaches consist in using SRS data alone or in combination with in situ data, usually streamflow data (Immerzeel & Droogers, 2008; Li et al., 2018; Rajib, Evenson, et al., 2018; Wambura et al., 2018). Calibration of hydrological models without concomitant streamflow data remains challenging, and attempts to do so have only shown limited success (Nijzink et al., 2018; Silvestro et al., 2015; Sutanudjaja et al., 2014; Wanders et al., 2014).

The simultaneous calibration of hydrological models with streamflow and different combinations of complementary data from SRS is increasingly discussed in recent literature (Stisen et al., 2018). Multivariate (i.e., multiple variables) parameter estimation (Efstratiadis & Koutsoyiannis, 2010) can substantially reduce the feasible model and parameter space and lead to more realistic internal model dynamics and related hydrological signatures (Clark et al., 2017; Shafii & Tolson, 2015), which can ultimately enhance the overall representation of catchment functioning (Bergström et al., 2002; Rakovec et al., 2016). Furthermore, and intimately linked to the above, multivariate calibration strategies can considerably reduce equifinality (i.e., nonidentifiable model parameters in inverse modeling approaches; Beven, 2006a; Savenije, 2001) and reduce prediction uncertainty (Fenicia et al., 2008; Fovet et al., 2015; Gupta et al., 1998; Gupta et al., 2008; Hrachowitz et al., 2014; Schoups et al., 2005). However, important open questions remain with respect to the combination of SRS data with streamflow data for model parameter estimation. While some studies observed a significant improvement in the representation of model outputs after SRS data incorporation (Chen et al., 2017; Leroux et al., 2016; Werth et al., 2009; Yassin et al., 2017), others found minor changes or even major deteriorations (Stisen et al., 2018; Tangdamrongsub et al., 2017; Tobin & Bennett, 2017). Such apparently contradictory conclusions are case study specific and need to be understood as resulting from model structures, model parametrizations, and trade-offs between improving water balance estimates and related streamflow dynamics and better representing other hydrological fluxes and states (Euser et al., 2013; Koppa et al., 2019;

Yassin et al., 2017). More generally, the key challenge results from the integration of several data sources (SRS or in situ) in parameter estimation, which can be attributed to conflicting information from different types of SRS data. Nonetheless, multivariate parameter estimation with SRS data remains promising, especially when streamflow data availability is limited or the data quality is questionable.

Although SRS data are more accessible with higher spatiotemporal resolution compared to in situ observations, they are generally not direct measurements of hydrological processes, which adds a level of uncertainty to any SRS-based parameter estimation study (Ehlers et al., 2018; Knoche et al., 2014; Ma et al., 2018). However, they provide spatial information on hydrological processes, which makes them a unique and relevant information source for spatially distributed representations of the system in models (Stisen et al., 2018; Wambura et al., 2018). For instance, many studies report different model performances when using different satellite-based products as input (e.g., precipiation; Pomeon et al., 2018; Thiemig et al., 2013) or as calibration variables (e.g., evaporation, soil moisture, and terrestrial water storage; Bai, Liu, & Liu, 2018; Nijzink et al., 2018). Nevertheless, for a given region, different products can give considerably different absolute values of a specific variable while they may exhibit plausible and similar spatial patterns (Beck et al., 2017; Dembele & Zwart, 2016). Additionally, retaining only the spatial pattern information of SRS data can substantially mitigate the uncertainty resulting from the fact that they are not direct observations, as long as their relative values are used rather than their absolute values (Mendiguren et al., 2017; Wambura et al., 2018).

In the context of using SRS data for DHM calibration, the simultaneous use of more than one SRS product to constrain several hydrological state or flux variables is uncommon (Clark et al., 2017; Lopez et al., 2017; Nijzink et al., 2018), as is the incorporation of spatial pattern in the calibration scheme using bias-insensitive metrics (Demirel et al., 2018; Zink et al., 2018). Using different variables from SRS products simultaneously in parameter estimation is in general not straightforward (Rajib et al., 2018; Silvestro et al., 2015; Tian et al., 2017) because they all have limitations (e.g., spatiotemporal resolutions and accuracy), which can lead to significant trade-offs in multivariate calibration (Koppa et al., 2019).

In light of the above, we propose to test a novel multivariate calibration strategy in which a DHM will be trained to simultaneously reproduce spatial patterns, i.e., relative spatial differences, of three variables from different SRS products describing different components of the hydrological system (i.e., evaporation, soil moisture, and terrestrial water storage), as well as in situ observations of streamflow. The proposed calibration framework combines simultaneously four noncommensurable variables and a new bias-insensitive metric for spatial pattern representation, which as a whole is different from previous studies (e.g., Demirel et al., 2018; Koppa et al., 2019; Nijzink et al., 2018; Rakovec et al., 2016; Zink et al., 2018) and therefore makes the novelty of this study. The following research hypotheses are tested:

1. Building upon previous work (e.g., Demirel et al., 2018; Rakovec et al., 2016; Zink et al., 2018), we assume that simultaneously calibrating a DHM on four noncommensurable variables and spatial patterns of satellite data considerably improves the predictive skill of the model, even for a DHM integrating a MPR scheme.
2. Our new bias-insensitive metric based on pixel-by-pixel locational matching can be used to improve the calibration of a DHM on observed spatial patterns of hydrological processes even in the presence of strong climatic gradients.

The overall goal of this study is to improve the spatial representation of dominant hydrological processes of a DHM without significantly deteriorating the streamflow signal and reproducing plausible dynamics of the hydrological system using spatial pattern information from SRS data sets. Such improvement will be an asset for spatial hydrology and large-domain water management applications (e.g., water accounting, drought monitoring, and flood prediction) and might subsequently lead to advances in prediction in ungauged basins (Blöschl et al., 2013; Hrachowitz et al., 2013; Sivapalan, 2003) with the use of readily accessible SRS data (Butler, 2014; Wulder & Coops, 2014). This work embraces the fourth paradigm for hydrology (i.e., data-intensive science, Peters-Lidard et al., 2017) and contributes to solving some of the issues (e.g., spatial variability and modeling methods) recently identified as the 23 unsolved problems in hydrology in the 21st century (Blöschl et al., 2019). The proposed multivariate calibration framework is tested with the mesoscale Hydrologic Model (mHM), with a case study in the poorly gauged Volta River basin (VRB) in West Africa.

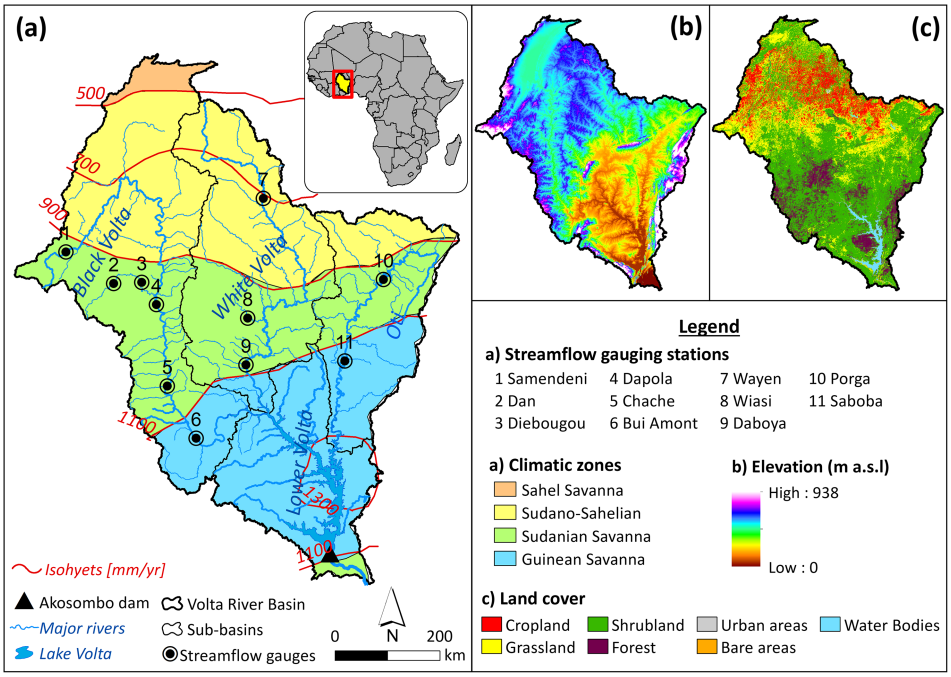

Figure 1. Physical and hydroclimatic characteristics of the Volta River basin.

2. Study Area

The transboundary VRB is the study area. It covers approximately 415,600 km^2 across six countries of West Africa. Figure 1 shows the physical and hydroclimatic characteristics of the VRB. The climate is characterized by a south-north gradient of increasing aridity and varies from subhumid in the south to semiarid in the north (Dembélé et al., 2019). Climate is driven by the Intertropical Convergence Zone, and four ecoclimatic zones (i.e., Sahelian, Sudano-Sahelian, Sudanian, and Guinean) can be identified (Figure 1a) based on the average annual precipitation and agricultural features (Food and Agriculture Organization/Global Information and Early Warning System, 1998; Mul et al., 2015). The characteristics of the four ecoclimatic zones are given in Table 1. Actual evaporation exceeds 80% of annual rainfall in the basin (Andreini et al., 2000; De Condappa & Lemoalle, 2009).

The topography is predominately flat as 95% of the relief is below 400 m above sea level (Figure 1b). The drainage system is composed of four sub-basins known as Black Volta (152,800 km^2), White Volta (113,400 km^2), Oti (74,500 km^2), and Lower Volta (74,900 km^2). The Volta River flows over 1,850 km and transits in the Lake Volta formed by the Akosombo dam before draining into the Atlantic Ocean at the Gulf of Guinea (Williams et al., 2016). Land cover (Figure 1c) is dominated by savannah formed by grassland interspersed with shrubs and trees covering about 88% of the basin area. Other land cover types include forest (9%), water bodies (2%), and bare land and settlements (1%).

Table 1
Characteristics of the Four Ecoclimatic Zones in the Volta River Basin

Ecoclimatic zones	Climate class	*AI* (−)	*P* (mm/year)	T_{avg} (°C)	T_{min} (°C)	T_{max} (°C)
Sahel Savanna	Arid	0.16 [0.12–0.20]	570 [470–610]	29 [29–30]	20 [20–21]	37 [36–38]
Sudano-Sahelian	Semiarid	0.29 [0.16–0.43]	790 [570–980]	29 [28–29]	20 [20–21]	36 [35–37]
Sudanian Savanna	Semiarid/dry subhumid	0.47 [0.33–0.98]	1,010 [890–1290]	28 [26–29]	21 [19–23]	35 [32–36]
Guinean Savanna	Dry subhumid/humid	0.70 [0.49–1.22]	1,190 [1030–1420]	28 [26–29]	21 [19–22]	34 [31–36]

Note. The annual mean value (with min-max range in brackets) is given for each variable. The aridity index (United Nations Environment Programme, 1997) is obtained from the global aridity index database (Trabucco & Zomer, 2018), and the WFDEI data (Weedon et al., 2014) are used for the long-term (1979–2016) estimation of annual precipitation and air temperature. *AI* = aridity index; *P* = precipitation, T_{avg} = average air temperature; T_{min} = minimum air temperature; T_{max} = maximum air temperature.

Table 2
Overview of the Modeling Data Sets

Variables	Product	Spatial resolution	Temporal resolution	Reference
Model setup				
Meteorological data				
Rainfall	CHIRPS v2.0	0.05°	Daily	Funk et al. (2015) http://chg.geog.ucsb.edu/data/chirps/
Temperature (average, minimum, and maximum)	WFDEI	0.5°	Hourly	Weedon et al. (2014) http://www.eu-watch.org/data_availability
Morphological data				
Terrain characteristics (elevation, slope, aspect, flow direction, and flow accumulation)	GMTED 2010	225 m (0.0021°)	Static	Danielson and Gesch (2011) https://topotools.cr.usgs.gov/
Soil properties (horizon depth, bulk density, and sand and clay content)	SoilGrids	250 m (0.0023°)	Static	Hengl et al. (2017) https://www.isric.org/explore/soilgrids
Geology	GLiM v1.0	0.5°	Static	Hartmann and Moosdorf (2012) https://doi.pangaea.de/10.1594/PANGAEA.788537
Land use land cover	Globcover 2009	300 m (0.0028°)	Static	Bontemps et al. (2011) http://due.esrin.esa.int/page_globcover.php
Phenology (leaf area index)	GIMMS	8 km (0.0833°)	Bimonthly	Tucker et al. (2005), Zhu et al. (2013) http://cliveg.bu.edu/modismisr/lai3g-fpar3g.html
Model calibration/evaluation				
In situ data				
Streamflow	-	Point	Daily	Multiple organizations (see the Acknowledgements section)
Complementary satellite products				
Terrestrial water storage anomaly	GRACE TellUS v5.0	1°	Monthly	Tapley et al. (2004), Landerer and Swenson (2012) https://grace.jpl.nasa.gov/
Surface soil moisture	ESA CCI SM v4.2	0.25°	Daily	Dorigo et al. (2017) https://www.esa-soilmoisture-cci.org/
Actual evaporation	GLEAM v3.2a	0.25°	Daily	Martens et al. (2017), Miralles et al. (2011) https://www.gleam.eu/
Land surface temperature (only for model evaluation)	MYD11A2 v6	1 km (0.0083°)	8-day	Wan et al. (2015) https://lpdaac.usgs.gov/products/myd11a2v006/

Note. CHIRPS = Climate Hazards Group InfraRed Precipitation with Station data; ESA CCI SM = European Space Agency Climate Change Initiative soil moisture; GIMMS = Global Inventory Modelling and Mapping Studies; GLEAM = Global Land Evaporation Amsterdam Model; GLiM = Global Lithological Map; GMTED = Global Multi-resolution Terrain Elevation Data; GRACE = Gravity Recovery and Climate Experiment; WFDEI = WATCH Forcing Data methodology applied to ERA-Interim data.

3. Data Sets

The data sets used to set up and run the distributed model for the 2000–2012 period include the basin morphological data (elevation, slope, land cover, etc.) and meteorological data (i.e., rainfall and air temperature). In situ streamflow data and complementary data from SRS are used to calibrate and to evaluate the model performance. A description of the data sets with their characteristics and their sources is given in Table 2.

The streamflow data were obtained from different organizations (see the Acknowledgements section). Previous work by Dembélé et al. (2019) describes the preprocessing of the streamflow data set, which was quality checked and whose missing data portions were gap filled.

Concerning the SRS products, the terrestrial water storage (S_t) anomaly data derived from changes in surface mass, which is related to the Earth's gravity field, is obtained from the Gravity Recovery and Climate Experiment (GRACE; Landerer & Swenson, 2012; Tapley et al., 2004). Over land, S_t is the sum of snow, ice, surface water, soil moisture, and groundwater. The data release RL05 (Swenson, 2012) is used in this study. It is a simple arithmetic mean of different solutions from three processing centers: Jet Propulsion Laboratory, Center for Space Research at University of Texas, and Geoforschungs Zentrum Potsdam.

Sakumura et al. (2014) found this ensemble mean product more effective in reducing noise in the Earth's gravity signal compared to the individual products. As the original baseline for GRACE-derived S_t anomaly data is the period 2004–2009, the S_t data is converted to a new baseline corresponding to the modeling period (2003–2012) used in this study, by averaging each grid point over the new baseline and subtracting that value from all time steps (National Aeronautics and Space Administration, 2019).

The surface soil moisture (S_u) data for a soil layer depth of 2–5 cm is obtained from European Space Agency Climate Change Initiative (ESA CCI; Dorigo et al., 2017). The combined product used in this study is a blended product of both active and passive microwave products derived from scatterometer (European Remote-Sensing Satellite (ERS), Active Microwave Instrument (AMI) and Advanced Scatterometer (ASCAT)) and radiometer (Scanning Multichannel Microwave Radiometer (SMMR), Special Sensor Microwave/Imager (SSM/I), Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI), Advanced Microwave Scanning Radiometer - Earth Observing System (AMSR-E), WindSat, Advanced Microwave Scanning Radiometer 2 (AMSR2), and Soil Moisture and Ocean Salinity (SMOS)) retrievals (Liu et al., 2012; Wagner et al., 2012). The merging algorithm of the combined product version 4.2 is described by Gruber et al. (2017).

Actual evaporation (E_a) data are obtained from the Global Land Evaporation Amsterdam Model (GLEAM) land surface model that uses satellite data as input (Martens et al., 2017; Miralles et al., 2011). It separately estimates the components of terrestrial evaporation (i.e., transpiration, bare soil evaporation, open-water evaporation, interception loss, and sublimation) based on the fraction of land cover types (i.e., bare soil, low vegetation, tall vegetation, and open water) before aggregating them for each grid cell. In GLEAM, potential evaporation (E_p) is calculated based on the Priestley and Taylor (1972) equation and thereafter converted into transpiration or bare soil evaporation using a stress factor, which is a parameter that accounts for environmental conditions limiting evaporation. The stress factor is estimated from microwave vegetation optical depth (i.e., water content in vegetation) and root-zone soil moisture that is calculated with a multilayer water balance algorithm. The fraction of open-water evaporation is assumed to equal E_p. The Gash (1979) analytical model further refined by Valente et al. (1997) is used to calculate rainfall interception by forests.

Land surface temperature (T_s) data from the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument of the National Aeronautics and Space Administration satellites are used as independent data for model evaluation, and they are not used during model calibration. The daytime product from the Aqua platform is used because that satellite passes over our study region around 13:45 in local time, corresponding to the highest daily temperature period with a clear-sky coverage (Wan et al., 2015).

For a full description of the data sets, the reader is referred to the corresponding data references in Table 2.

4. Experimental Design

4.1. Hydrological Model

mHM (version 5.8) is a spatially explicit (i.e., fully distributed) conceptual model based on numerical approximations of dominant hydrological processes per grid cell in the modeling domain (Kumar et al., 2013; Samaniego et al., 2010). The following processes can be represented: canopy interception, snow accumulation and melting, soil moisture dynamics, infiltration and surface runoff, evaporation, subsurface storage and discharge generation, deep percolation and baseflow, and discharge attenuation and flood routing. The total grid-generated runoff is routed to the neighboring downstream cell following the river network using a multiscale routing model (Thober et al., 2019) based on the Muskingum-Cunge method (Cunge, 1969). mHM uses an MPR technique (Samaniego et al., 2017) to account for the subgrid variability of the basin physical characteristics (e.g., topography, soil texture, geology, and land cover properties). Pedotransfer functions and global parameters are used to link the model parameters (e.g., hydraulic conductivity and soil porosity) to the basin physical characteristics. There are at most 53 global or super parameters (Pokhrel et al., 2008), which are time- and space-invariant parameters tuned during the calibration procedure. The model uses three different scales of spatial discretization (i.e., grid cell resolution) of the modeling domain corresponding to operation levels. The first scale is used for morphologic data (L0), the second is dedicated to fluxes and states calculation (L1), and the third scale is considered for the meteorological data (L2). L1 should be a multiple of L0 and a submultiple of L2. Three land use and land cover classes are

considered by mHM: forests (e.g., coniferous and deciduous), permeable areas (e.g., grasslands, croplands, and bare soils), and impervious areas (e.g., urban and built-up areas, water bodies, and consolidated soils).

4.2. Model Implementation

For the current setup of the mHM model, reference evaporation (E_{ref}) is calculated based on the Hargreaves and Samani (1985) method, which only requires temperature data. Following Demirel et al. (2018), a dynamically scaling function (F_{DS}) is used to calculate potential evaporation (E_p) to account for vegetation-climate interactions and improve spatial estimation of E_p (Bai et al., 2018; Jiao et al., 2017). The approach is based on the concept of E_p calculation from crop coefficients and E_{ref} (Allen et al., 1998; Birhanu et al., 2019). Here the F_{DS} (equation (1)) plays the role of a spatially varying crop coefficient, and it is estimated based on leaf area index (I_{LA}) data. It is formulated as follows:

$$F_{DS} = a + b\left(1 - e^{(c \cdot I_{LA})}\right), \tag{1}$$

where a is the intercept term representing uniform scaling, b represents the vegetation-dependent component, and c describes the degree of nonlinearity in the I_{LA} dependency. The coefficients a, b, and c are determined during model calibration.

$$E_p = F_{DS} \times E_{ref}. \tag{2}$$

Total actual evaporation (i.e., all forms of evaporation including transpiration, E_a) is estimated as a fraction of E_p from soil layers and depends on the fraction of vegetation roots and soil moisture availability (Feddes et al., 1976). Soil moisture is calculated following a multilayer infiltration capacity approach using a three-layer soil scheme. The depths of the soil layers are 5, 30, and 100 cm. Terrestrial water storage in mHM is calculated per grid cell by summing up all the subsurface water storage (i.e., reservoirs generating soil moisture, interflow, and baseflow) and the surface water storage in sealed areas. There is no snowfall in the VRB. In this study, 36 global parameters are determined through model calibration (see supporting information Table S2). The full description of the model and the formulation of the hydrological processes can be found in the work of Samaniego et al. (2010). All morphological data are resampled to a resolution of 1/512° (~200 m at the equator) and the meteorological data to 0.0625° (~7 km). The nearest neighbor technique is used to resample categorical data, while bilinear resampling is used for continuous data. The model is run at daily time step with a spatial resolution of 0.25° (~28 km), corresponding to 619 modeling grid cells in the basin, but taking into account the subgrid variability of the morphological data using the MPR technique (Samaniego et al., 2010; Samaniego et al., 2017). Every modeling grid cell (L1) contains 16 meteorological grid cells (L2) and 16,384 morphological grid cells (L0).

4.3. Model Calibration and Evaluation Strategies

The modeling period spans from 2000 to 2012 and consists of 3 years (2000–2002) model warm-up period, 6 years (2003–2008) calibration period, and 4 years (2009–2012) evaluation period. Based on data availability and quality in the VRB (Dembélé et al., 2019), 11 streamflow gauging stations are chosen to have a good coverage of the river network (Figure 1), and the calibration is done on them simultaneously to obtain a single-parameter set for the whole VRB. The domain-wide calibration, which was proven to give similar performance as the domain-split calibration (Mizukami et al., 2017), is preferred here because of the limited number of streamflow stations and for seamless spatial pattern representation across sub-basins (see Figure 1).

Two main calibration approaches are adopted to evaluate the benefit of including spatial patterns in multivariate parameter estimation with SRS data. The first approach is the streamflow-only calibration, and the second approach uses multiple SRS data sets in addition to streamflow. In both cases, the formulation of the objective functions follows the Euclidian distance approach in which all elements are equally weighted (Khu & Madsen, 2005); see equation (A1).

4.3.1. Calibration on Streamflow—Benchmark

The first calibration approach is the benchmark calibration case (case Q) where the hydrological model is constrained with in situ streamflow (Q) data only. An objective function Φ_Q (equation (3)) combines the Nash-Sutcliffe efficiency (Nash & Sutcliffe, 1970) of streamflow (E_{NS}) and the Nash-Sutcliffe efficiency of

the logarithm of streamflow (E_{NSlog}) presented in equations (A2) and (A3), and it is formulated as Euclidean distance such that it has to be minimized:

$$\Phi_Q = \frac{1}{g}\sum_1^g \sqrt{(1-E_{NS})^2 + (1-E_{NSlog})^2}, \tag{3}$$

where g is the number of streamflow gauging stations present within the modeling domain. Φ_Q is obtained by equally weighing the streamflow gauging stations, and it ranges from 0, its ideal value, to $+\infty$.

4.3.2. Calibration on Multiple Variables With Spatial Patterns

4.3.2.1. Spatial Pattern Efficiency Metric

The degree of reproduction of the spatial patterns of E_a and S_u is quantified with a proposed pattern matching metric, denoted E_{SP}. The development of E_{SP} is motivated by the need for simplicity and robustness in pattern matching with respect to existing metrics (cf. Koch et al., 2015; Koch et al., 2018). It is a multicomponent metric formulated in a way functionally equivalent to the Kling-Gupta efficiency (cf. equation (A6); Gupta et al., 2009; Kling et al., 2012) and the SPAtial EFficiency metric (*SPAEF*, Demirel et al., 2018; Koch et al., 2018). It simultaneously assesses the matching of the spatial distribution of grid cells, the relative variation, and the strength of the monotonic relationship between the observed and estimated variables. Therefore, E_{SP} is a bias-insensitive metric that focuses on the patterns of the variables rather than their magnitudes. Considering a modeled variable (X_{mod}) and an observed variable (X_{obs}) of n elements, E_{SP} is defined as follows:

$$E_{SP} = 1-\sqrt{(r_s-1)^2 + (\gamma-1)^2 + (\alpha-1)^2}, \text{with} \tag{4}$$

$$r_s = 1-\frac{6\sum_1^n d^2}{n(n^2-1)}, \tag{5}$$

$$\gamma = \frac{\frac{\sigma_{mod}}{\mu_{mod}}}{\frac{\sigma_{obs}}{\mu_{obs}}}, \text{and} \tag{6}$$

$$\alpha = 1-E_{RMS}(Z_{X_{mod}}, Z_{X_{obs}}), \tag{7}$$

where r_s is the Spearman rank-order correlation coefficient with d the difference between the ranks of X_{mod} and X_{obs}. γ is the variability ratio (i.e., ratio of coefficients of variation) that assesses the similarity in the dispersion of the probability distributions of X_{mod} and X_{obs}, with μ and σ representing the mean and the standard deviation, respectively, and α the spatial location matching term calculated as the root-mean-square error (E_{RMS}) of the standardized values (z-scores, Z_X) of X_{mod} and X_{obs} (see equations (A4) and (A5)). The z-score is a standardization of the scale of a distribution that facilitates its comparison with another distribution. The z-scores identify and describe the exact location of each observation in a distribution (Gravetter & Wallnau, 2013). For a given variable with values represented spatially as a 2-D matrix, the z-scores represent the number of standard deviations the value in each grid cell is from the population mean (Oyana & Margai, 2015). Consequently, forcing the z-scores of X_{mod} and X_{obs} to be equal (i.e., minimizing their E_{RMS}) corresponds to matching their grid cell locations (i.e., spatial patterns). Finally, E_{SP} is formulated such that it ranges from $-\infty$ to 1, which is its optimal value. Contrarily to *SPAEF*, E_{SP} does not require any user-defined parameter (i.e., number of bins in *SPAEF*), and it uses a nonparametric correlation coefficient (i.e., r_s), which limits its sensitivity to outliers as opposed to the Pearson correlation coefficient (Legates & McCabe, 1999; Pool et al., 2018; Spearman, 1904) used in *SPAEF*. A comparison of E_{SP} to *SPAEF* is provided in the supporting information (Figure S40 and Table S3).

4.3.2.2. Multivariate Calibration Strategies

In contrast to the first calibration strategy, which only considers Q as target variable (case Q), the second calibration strategy involves multiple variables (case MV). The potential improvement of the modeled fluxes and states is estimated by constraining the parameter estimation with a simultaneous combination of three variables from SRS products (S_t, S_u, and E_a), in addition to Q. Here the spatial patterns of E_a and S_u are used in the multivariate calibration, while the temporal dynamics of S_t are averaged over the whole basin due to

Table 3
Variants of Multivariate Calibration Cases in the Leave-One-Out Approach

Calibration case	Calibration variable	Objective function	Specificity	Equation number
Case MV-E_a	Q, S_t, S_u	$\Phi_{MV-E_a} = \sqrt{\Phi_Q{}^2 + \Phi_{S_t}{}^2 + \Phi_{S_u}{}^2}$	No direct constraint on evaporation	(12)
Case MV-S_u	Q, S_t, E_a	$\Phi_{MV-S_u} = \sqrt{\Phi_Q{}^2 + \Phi_{S_t}{}^2 + \Phi_{E_a}{}^2}$	No specific constraint on surface soil moisture	(13)
Case MV-S_t	Q, S_u, E_a	$\Phi_{MV-S_t} = \sqrt{\Phi_Q{}^2 + \Phi_{S_u}{}^2 + \Phi_{E_a}{}^2}$	No direct constraint on deep subsurface processes	(14)
Case MV-Q	S_t, S_u, E_a	$\Phi_{MV-Q} = \sqrt{\Phi_{S_t}{}^2 + \Phi_{S_u}{}^2 + \Phi_{E_a}{}^2}$	Only satellite-based variables with no direct constraint on streamflow	(15)

the relatively coarse spatial resolution of the GRACE data. The temporal dynamics of S_t per grid cell is also assessed during model evaluation. The multivariate objective function Φ_{MV} (equation (8)) is defined as follows:

$$\Phi_{MV} = \sqrt{\Phi_Q{}^2 + \Phi_{S_t}{}^2 + \Phi_{S_u}{}^2 + \Phi_{E_a}{}^2}, \text{with} \tag{8}$$

$$\Phi_{E_a} = 1 - \frac{1}{t}\sum_1^t E_{SP}\left(E_{a,mod}(t), E_{a,obs}(t)\right), \tag{9}$$

$$\Phi_{S_u} = 1 - \frac{1}{t}\sum_1^t E_{SP}\left(S_{u,mod}(t), S_{u,obs}(t)\right), \text{and} \tag{10}$$

$$\Phi_{S_t} = E_{RMS}\left(Z_{S_{t,mod}}(t), Z_{S_{t,obs}}(t)\right), \tag{11}$$

where t is the number of time steps of the calibration period. The subobjective functions Φ_{E_a} and Φ_{S_u} are based on the E_{SP} (equation (4)) of modeled and observed E_a and S_u, while Φ_{S_t} denotes the root-mean-square error (E_{RMS}) of the z-scores of the modeled and observed basin-averaged S_t anomaly.

Consequently, Φ_Q ensures a reliable prediction of streamflow signatures (i.e., high and low flows), Φ_{E_a} and Φ_{S_u} serve to improve the spatial patterns of the modeled E_a and S_u, and Φ_{S_t} constrains the temporal dynamics of the modeled S_t, which should contribute to a better prediction of the water balance at monthly and annual scales. In fact, Φ_{E_a} and Φ_{S_u} are calculated such that the spatial pattern efficiencies of E_a and S_u are determined over the grid cells at each monthly time step, before averaging them over the calibration period, while Φ_{S_t} is calculated for the basin-averaged S_t over the calibration period. Note that in the E_{RMS} metric (equation (A5)), n denotes the number of grid cells in the spatial domain when calculating Φ_{E_a} and Φ_{S_u} (i.e., α in E_{SP}), while it corresponds to the length of the calibration period (i.e., $n = t$) in the calculation of Φ_{S_t}. All constituents of Φ_{MV} (i.e., $\Phi_Q, \Phi_{E_a}, \Phi_{S_u}$, and Φ_{S_t}) vary in the same range from 0 to $\pm\infty$. Therefore, Φ_{MV} has the same range of values with an optimal value of 0. The dynamically dimensioned search algorithm (Tolson & Shoemaker, 2007) is used for parameter estimation using 3,000 iterations. Daily streamflow data are used for model calibration, while monthly SRS data are preferred to avoid errors due to potential time lags among satellite sensor measurements and model simulations.

4.3.2.3. Contribution of Individual Variables to Multivariate Calibration

Assessing the individual contribution of variables used in multivariate calibration is rarely done, although it can help quantify trade-offs in modeling flux and states variables (Koppa et al., 2019). Here the contribution of each SRS data set to the multivariate calibration case is investigated with a leave-one-out approach. The procedure consists in removing one SRS data type from the calibration case MV and evaluating the predictive skill of the model. In addition, multivariate calibration without streamflow data, and thus exclusively based on SRS data, is tested to determine the potential of SRS data for hydrological model calibration in regions where little or no streamflow data are available. Consequently, four additional objective functions (Table 3) are used for multivariate calibration cases without E_a (case MV-E_a), without S_u (case MV-S_u), without S_t (case MV-S_t), and without streamflow (case MV-Q).

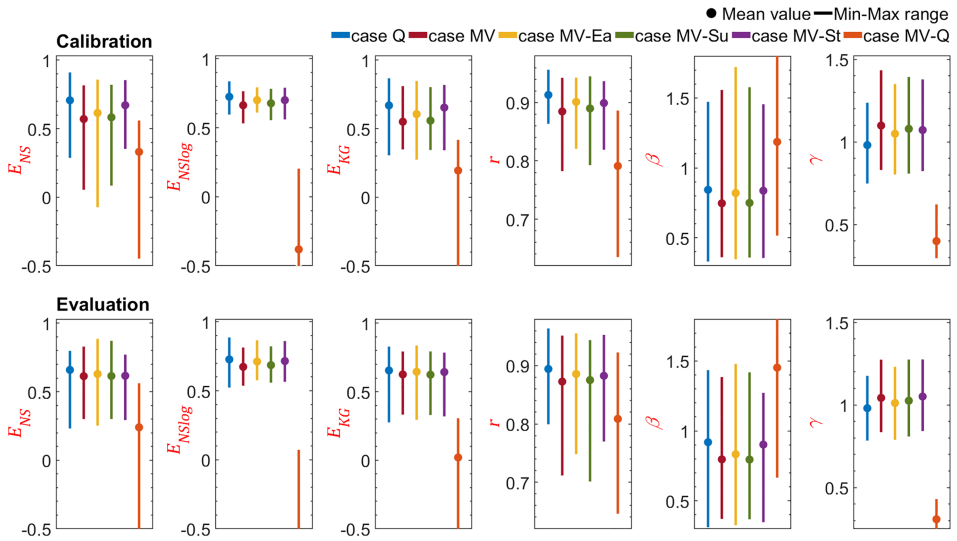

Figure 2. Statistics of model performance for streamflow. The best score of all the metrics is 1. The dots give the mean score, and the bars represent the min-max range of values for 11 streamflow gauges. The colors correspond to the model calibration cases.

4.4. Postcalibration Model Evaluation

The predictive skill of the model is evaluated by assessing the transferability of the global parameters across temporal periods and spatial scales obtained by the abovementioned calibration strategies. First, the temporal transferability is evaluated following a split-sample test that consists in assessing performances for a period that is different from the calibration period (Klemes, 1986). Secondly, spatial scale transferability is evaluated by using different grid cell (i.e., pixel) sizes as modeling resolution (Kumar et al., 2013; Samaniego et al., 2010). The global parameters of the model for all calibration cases are obtained for a resolution of 0.25° (~28 km, i.e., 619 pixels in the basin), and the same parameters are used to run the model without recalibration at four different finer scales: 0.125° (~14 km, i.e., 2,320 pixels), 0.0625° (~7 km, i.e., 8,974 pixels), 0.03125° (~3.5 km, i.e., 35,231 pixels), and 0.015625° (~1.75 km, i.e., 139,494 pixels). The evaluation data for model parameter transferability are streamflow for streamflow, using the Kling-Gupta efficiency (E_{KG}), and fine-scale T_s data to evaluate E_a and S_u, using r_s, while no high-resolution data are available for S_t evaluation. T_s is used as proxy data for the evaluation of S_u and E_a because past studies found significant negative correlation between T_s and S_u (Kumar et al., 2013; Lakshmi et al., 2003; Wang et al., 2007) and a control of T_s over E_a (Boni et al., 2001; Lakshmi, 2000).

Following Biondi et al. (2012), supplemental skill metrics different from those used in model calibration are computed for a thorough model evaluation because every metric has its own limitations (Fowler et al., 2018; Knoben et al., 2019; Santos et al., 2018; Schaefli & Gupta, 2007). In addition to E_{NS}, E_{NSlog}, E_{SP}, E_{RMS}, and r_s, the E_{KG} is reported for model evaluation.

5. Results and Discussions

The following section presents and discusses the results of model performances for different variables used in the calibration procedure. The results refer to the evaluation period when analyzing the results if not clearly specified. However, both calibration and evaluation results are presented in figures. Hereafter, the SRS data sets are called reference data, as they are not direct observations. Detailed results on model performances for each of the four climatic zones in the VRB are provided as supporting information (Figures S17–S33).

5.1. Model Performance for Streamflow

The model performance for streamflow (Q) at the 11 gauging stations is given in Figure 2. For the calibration period, the mean E_{KG} is 0.67 ($E_{NS} = 0.71$, $E_{NSlog} = 0.72$) for the model calibration with only Q data (i.e., case

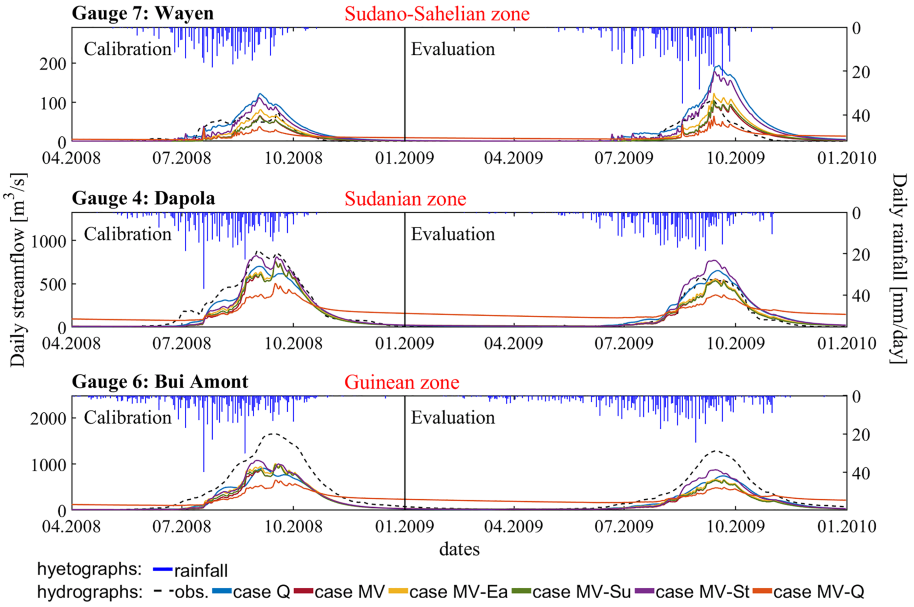

Figure 3. Hydrographs at selected stations in different climatic zones for all model calibration cases. Only a subset of the simulation period (2003–2012) is shown for visualization.

Q). The performance of Q in the calibration period decreases when multiple variables are used to constrain the parameter search. The mean E_{KG} is 0.55 ($E_{NS} = 0.57$, $E_{NSlog} = 0.66$) for the multivariate calibration (i.e., case MV), corresponding to a decrease of 18% compared to that in case Q.

Regarding the other multivariate calibration cases, the best performance with respect to Q is obtained in case MV-S_t with a mean E_{KG} of 0.65 ($E_{NS} = 0.67$, $E_{NSlog} = 0.70$), which represents a slight decrease of 3% compared to that in case Q, while the weakest performance is given by case MV-Q with a mean E_{KG} of 0.19 ($E_{NS} = 0.33$, $E_{NSlog} = -0.38$). Differences in measurement scales between Q data and satellite products (i.e., river dimensions vs. pixel size) can justify a low performance for case MV-Q (see section 5.5). In general, all calibration cases give a good timing of Q with a mean correlation coefficient of $r > 0.79$, but they underestimate it with a mean bias of $\beta < 0.85$, except case MV-Q, which shows overestimation with a positive bias ($\beta = 1.19$). They all show a higher variability of Q than the observed data, with mean $\gamma > 1.04$, except case Q ($\gamma = 0.98$) and case MV-Q ($\gamma = 0.40$), which produce a lower variability. A subset of the hydrographs of three stations from different climatic zones are depicted in Figure 3. More statistics on the model performance along with the complete hydrographs and flow signatures (i.e., flow duration curves and seasonal streamflow) are provided in the supporting information (Table S1 and Figures S1–S11).

During the evaluation period, as compared to the calibration period, the model performance for the mean E_{KG} decreases by 2% for case Q (from 0.67 to 0.66) and case MV-S_t (from 0.65 to 0.64) and 90% for case MV-Q (from 0.19 to 0.02), while it increases by 7% for case MV-E_a (from 0.61 to 0.65), 12% for case MV-S_u (from 0.56 to 0.62), and 14% for case MV (from 0.55 to 0.63). Considering the mean E_{KG}, case MV performs less well than case Q by 11% on average, which means 18% less during the calibration and 4% less during the evaluation period. The deterioration of streamflow performance in a multivariate calibration setting is also reported in previous studies (Bai, Liu, & Liu, 2018; Livneh & Lettenmaier, 2012; Pomeon et al., 2018; Rakovec et al., 2016). However, this is largely an artefact of Type I error (i.e., falsely accepting poor models; Beven, 2010) induced by the Q-only calibration, resulting in inconsistency in the representation of processes (Gupta et al., 2012; Hrachowitz et al., 2014). In addition, the performance of Q slightly increases when E_a (+7%) or S_t (+11%) are left out of the multivariate calibration with case MV during the evaluation period. Therefore, as shown in Figure 2, the combinations $Q + S_t + S_u$ (i.e., case MV-E_a) and $Q + S_u + E_a$ (i.e., case MV-S_t) are the best for streamflow prediction, while $Q + S_t + E_a$ (i.e., case MV-S_u) performs similar to $Q + S_t + S_u + E_a$ (i.e., case MV).

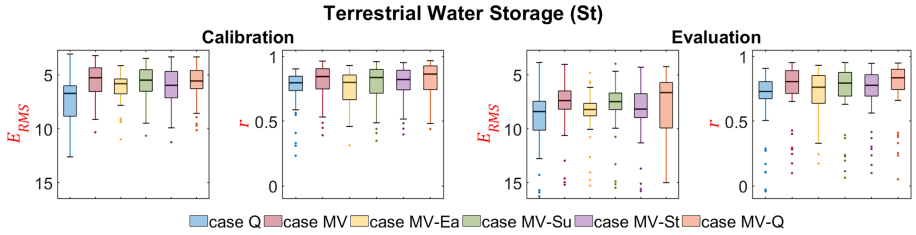

Figure 4. Statistics of model performance for terrestrial water storage (S_t). The y-axis is reversed for E_{RMS}. The number of elements per boxplot ($n = 52$) corresponds to the number of gird cells for GRACE data in the study area. The colors correspond to the model calibration cases.

5.2. Model Performance for Terrestrial Water Storage

The statistics for the monthly terrestrial water storage (S_t) anomalies are given in Figure 4. The results per climatic zone is provided in the supporting information (Figure S19). A similar trend in skill scores (i.e., E_{RMS} and r) among all calibration cases is observed in the calibration and evaluation periods with weaker scores during evaluation.

The evaluation period is characterized by a substantial improvement from model case Q (median $E_{RMS} = 8.41$ cm, $r = 0.73$) to case MV ($E_{RMS} = 7.38$ cm, $r = 0.81$). In general, all multivariate calibration cases reproduce the variability in S_t better than did case Q. Previous studies also reported improvement of S_t prediction in multivariate settings (Chen et al., 2017; Livneh & Lettenmaier, 2012; Werth et al., 2009). The lowest performance increase is observed when E_a ($E_{RMS} = 8.21$ cm, $r = 0.76$) or S_t ($E_{RMS} = 8.18$ cm, $r = 0.78$) is removed from the multivariate setting. The best prediction is obtained with case MV-Q, yielding median E_{RMS} of 6.65 cm and r of 0.84.

Figure 5a shows the climatology of the basin-averaged S_t for all models. The full monthly time series (Figure S17) and the climatological trends (Figure S18) per climatic zone are provided in the supporting information. The temporal dynamics of the normalized GRACE-derived S_t is well reproduced by all models ($r > 0.89$) with different degrees of underestimation from September to March, which is a period with little or no rainfall, and slight overestimation from April to June, which is the beginning of the rainy season. All models fit well the period July–August, which is the wettest period of the rainy season. In general, case MV-E_a shows the highest deviation from the satellite signal ($r = 0.90$) followed by case Q ($r = 0.92$). Removing spatial patterns of E_a from the calibration leads to an S_t overestimation during the rainy season and an underestimation during the dry season. The same trend can be observed when only Q is used for model calibration (i.e., case Q). The S_t simulation improves in the multivariate calibration including Q (i.e., case MV, $r = 0.97$), but the best match is obtained when Q is left out (i.e., case MV-Q, $r = 0.99$). When GRACE-derived S_t is excluded from the parameter estimation (i.e., case MV-S_t), the model still performs well for S_t climatology with $r = 0.96$. Consequently, E_a is the most critical variable for predicting the S_t signal in the proposed multivariate calibration setting, while S_u is less critical, probably because the GRACE-derived S_t signal already accounts for S_u (Li et al., 2012).

5.3. Model Performance for Soil Moisture

The climatology of the basin-averaged reference (i.e., ESA CCI) soil moisture (S_u) and different modeled S_u are depicted in Figure 5b. The maps of climatology (Figures S13 and S14), the full monthly time series (Figure S20), and the climatological trends (Figure S21) per climatic zone are provided in the supporting information. All calibration cases give a good performance ($r > 0.91$), with a good representation of S_u seasonality during both calibration and evaluation periods.

All simulations overestimate the reference S_u during the rising limb (February–August), corresponding to the increasing rainfall period, and underestimate it during the recession limb (September–January). Simulations that show the highest deviation from the reference during the rising limb, on the contrary, show the lowest deviation during the recession and vice versa. The overall best performance is obtained with case MV-E_a ($r = 0.96$), with a better match when the basin is not water limited (i.e., September–January). Case MV-S_t and case MV-S_u show similar performances ($r \approx 0.95$) with a consistent deviation from the

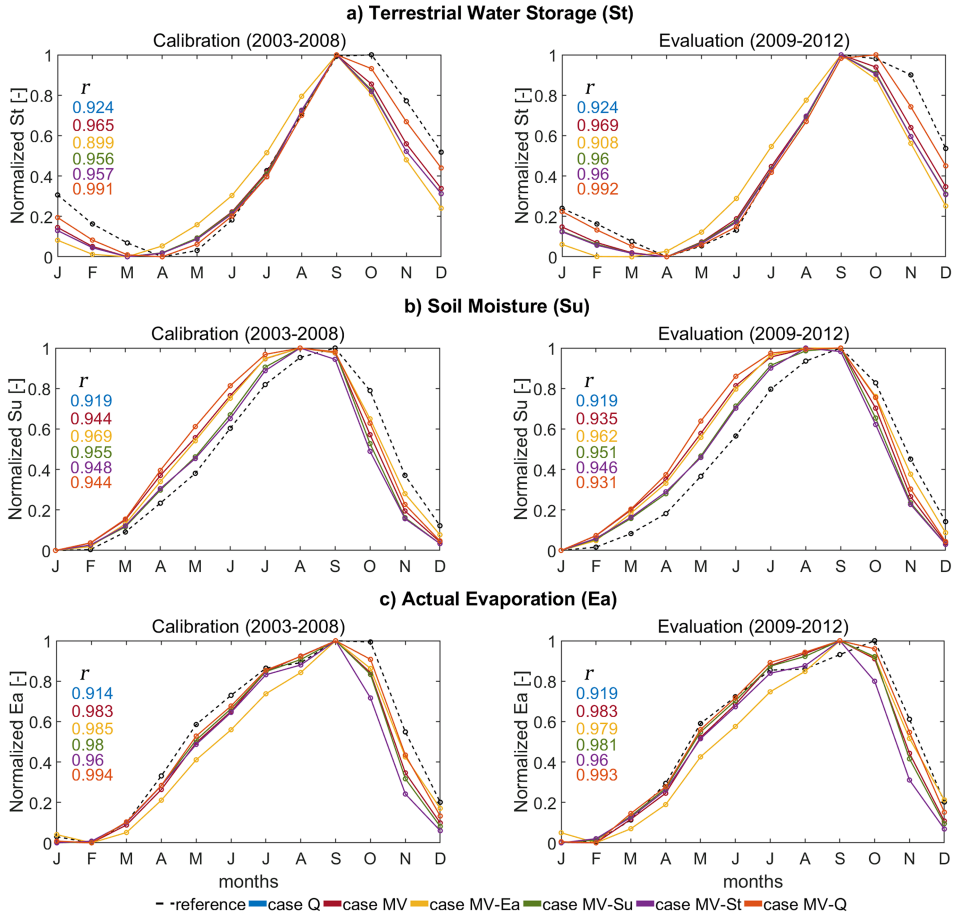

Figure 5. Climatology of (a) terrestrial water storage, (b) soil moisture, and (c) actual evaporation with the Pearson correlation coefficient (r) indicating the performance of all model calibration cases.

reference S_u for all months, while case MV and case MV-Q have similar performances ($r \approx 0.94$) but with a better fit of the reference S_u in the recession limb. It can be inferred that Q is the most critical variable for S_u reproduction during the rising limb, while S_t and satellite S_u improve the simulation during the recession limb. Case Q outperforms all multivariate calibration cases when soil water content increases, and it underperforms them when the maximum water content is reached and starts decreasing. The overall lowest performance is given by case Q ($r = 0.92$), followed by case MV-Q ($r = 0.93$), suggesting that Q alone is not sufficient for predicting the temporal dynamics of S_u, but it remains useful in the multivariate calibration setting. Surprisingly, E_a does not bring substantial information to the multivariate prediction of S_u. Contrastingly, Pomeon et al. (2018) obtained a slight improvement in S_u simulation (+7%) when using absolute values of satellite E_a in their multivariate calibration. The model performance in reproducing spatial patterns is measured with E_{SP}, and its components (i.e., r_s, γ, and α) are summarized in Figure 6a. The results for each month (Figure S41) and for the climatic zones (Figures S22–S25) are provided in the supporting information.

The evaluation and calibration periods depict similar trends when comparing the different metrics (i.e., E_{SP}, r_s, γ, and α). The lowest performance is given by the Q-only calibration (i.e., case Q) with median $E_{SP} = -0.07$ ($r_s = 0.54$, $\gamma = 0.85$, and $\alpha = 0.13$), followed by case MV-S_u with median $E_{SP} = -0.02$ ($r_s = 0.53$, $\gamma = 0.84$, and $\alpha = 0.16$). Case MV shows better performances with median $E_{SP} = 0.03$ ($r_s = 0.57$, $\gamma = 0.88$, and $\alpha = 0.19$). The best performance even with a low score is given by case MV-E_a with median $E_{SP} = 0.09$ ($r_s = 0.60$, $\gamma = 0.93$, and $\alpha = 0.19$), followed by case MV-Q with median $E_{SP} = 0.05$ ($r_s = 0.58$, $\gamma = 0.82$, and $\alpha = 0.18$) and case MV-S_t with median $E_{SP} = 0.02$ ($r_s = 0.55$, $\gamma = 0.87$, and $\alpha = 0.16$). Removing E_a or Q from the multivariate setting results in better performances, while the removal of S_t or satellite S_u results in lower performances.

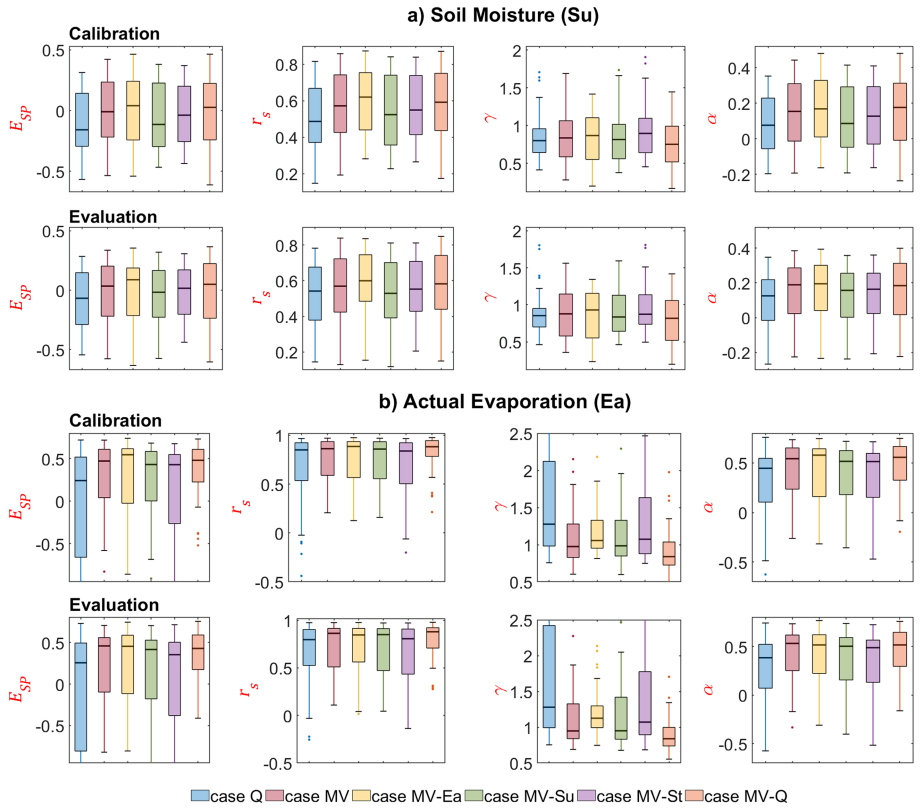

Figure 6. Spatial statistics of model performance for (a) soil moisture and (b) actual evaporation. The best score is 1 for all the metrics. The number of elements per boxplot corresponds to the number of months in the calibration period ($n = 72$) or in the evaluation period ($n = 48$). The colors correspond to the model calibration cases.

Consequently, satellites S_u and S_t are the most important variables for improving the spatial patterns of modeled S_u. Better simulation of S_u in multivariate settings is also reported by Lopez et al. (2017) using $E_a + S_u$ calibration and by Pomeon et al. (2018) with $Q + E_a$ calibration.

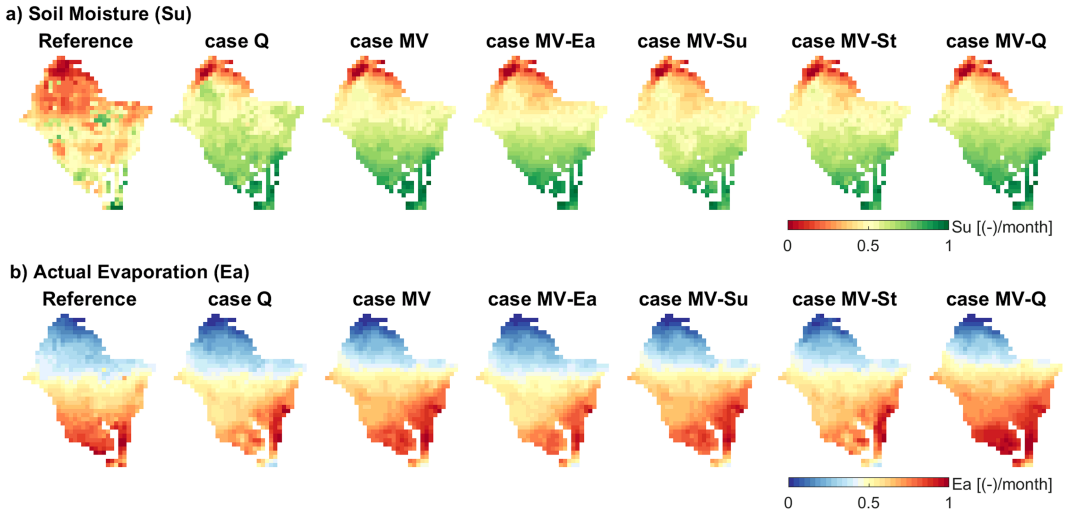

Figure 7. Long-term monthly average of (a) soil moisture and (b) actual evaporation for all model calibration cases over the simulation period (2003–2012). The reference map represents the satellite product (ESA CCI for S_u and GLEAM for E_a). Masked pixels are gaps in satellite measurements or lake areas not modeled in mHM. The values are normalized for better emphasizing on patterns and using a unique color scale.

The long-term monthly average (2003–2012) of S_u is illustrated in Figure 7a. See section 5.5.2 for S_u comparison with T_s. Although the spatial patterns of modeled S_u in the multivariate cases are still different from the reference S_u (Figure 7a), they are better than the Q-only case.

5.4. Model Performance for Evaporation

The climatology of basin-averaged reference (i.e., GLEAM) actual evaporation (E_a) and different modeled E_a are depicted in Figure 5c. The maps of climatology (Figures S15 and S16), the full monthly time series (Figure S26), and the climatological trends (Figure S27) per climatic zone are provided in the supporting information. All calibration cases give a good performance ($r > 0.91$), reproducing well E_a seasonality during both the calibration and evaluation periods.

The best performance is obtained with case MV-Q ($r = 0.99$). In general, all simulations tend to underestimate E_a. During the rising limb (February–August), corresponding to the increasing rainfall period, the highest deviation is given by case MV-E_a, although the overall performance is good ($r = 0.98$). Mismatches are more prominent during the recession limb (September–January), where all modeled E_a decrease faster than the reference. The highest deviation is observed when only Q data are used for model calibration ($r = 0.92$), followed by case MV-S_t ($r = 0.96$). It can be inferred that the model case MV-S_t is missing adequate information on the available water amount to be evaporated, which can be obtained from the satellite S_t signal. During the recession period, little to no rainfall occurs in the basin, but a part of antecedent rainfalls is stored in reservoirs and lakes, which represent a major source of land evaporation. It can be argued that Q alone is not sufficient for modeling E_a, while S_t brings additional and useful information for simulating E_a, which supports our research hypothesis. The performance of case MV-S_u is similar to that of case MV ($r \approx 0.98$), meaning that S_u is not critical for predicting the temporal dynamics of E_a. Moreover, satellite E_a improves the modeled E_a during water accumulation in the basin (i.e., February–August) and is no longer critical when the basin is not water limited (i.e., September–January). This result suggests that the model can mainly rely on GRACE-derived S_t to reproduce E_a. Similar results on the good estimation of E_a with GRACE-derived S_t are found in literature (e.g., Bai, Liu, & Liu, 2018; Livneh & Lettenmaier, 2012; Rakovec et al., 2016). Pomeon et al. (2018) also obtained a higher model performance for E_a in their multivariate setting (i.e., $Q + E_a$) with mHM in West Africa.

Figure 6b gives the spatial pattern efficiency of E_a for all model calibration cases. The results for each month (Figure S41) and for the climatic zones (Figures S28–S31) are provided in the supporting information. In general, the performance decreases from the calibration period to the evaluation period, and the modeled E_a with all model calibration cases has higher spatial pattern efficiency scores ($E_{SP} > 0.25$) compared to modeled S_u ($E_{SP} < 0.1$). All multivariate calibration cases outperform the Q-only calibration, giving the lowest performance with median $E_{SP} = 0.28$. The Q-only calibration gives a good spatial correlation ($r_s = 0.8$) but overestimates the variability ($\gamma = 1.28$) and struggles to match the spatial location of grid cells ($\alpha = 0.39$) of E_a. The best spatial pattern matching is given by case MV with median $E_{SP} = 0.46$ ($r_s = 0.86$, $\gamma = 0.95$, and $\alpha = 0.53$). Removing Q from the multivariate setting (i.e., case MV-Q) results in an underestimation of the spatial variability of E_a, with median $E_{SP} = 0.43$ ($r_s = 0.88$, $\gamma = 0.84$, and $\alpha = 0.52$). In contrast, the spatial variability of E_a is overestimated for case MV-E_a with median $E_{SP} = 0.45$ ($r_s = 0.85$, $\gamma = 1.13$, and $\alpha = 0.52$), while case MV-S_u yields a lower spatial location score with median $E_{SP} = 0.42$ ($r_s = 0.85$, $\gamma = 0.95$, and $\alpha = 0.50$). The spatial pattern performance of E_a is more sensitive to the removal of S_t, as shown by case MV-S_t with median $E_{SP} = 0.35$ ($r_s = 0.81$, $\gamma = 1.07$, and $\alpha = 0.49$). These results indicate that spatial patterns of S_u can improve the spatial patterns of E_a and S_t is critical for reproducing both the temporal and spatial dynamics of E_a. Demirel et al. (2018) similarly reported better spatial pattern performance for E_a when using a multivariate setting (i.e., $Q + E_a$) compared to that when using the Q-only calibration.

Figure 7b illustrates the long-term (2003–2012) monthly average of E_a. See section 5.5.2 for E_a comparison with T_s. The southern region of the basin, with a subhumid climate, is where the multivariate calibration cases show more differences in spatial patterns compared to case Q. Besides the south-north differences, it is interesting to see strong differences in the west-east variability of the spatial pattern. As the southern part is subhumid ($E_a \geq 70\%$), small variations in E_a are not well represented when the model is calibrated using

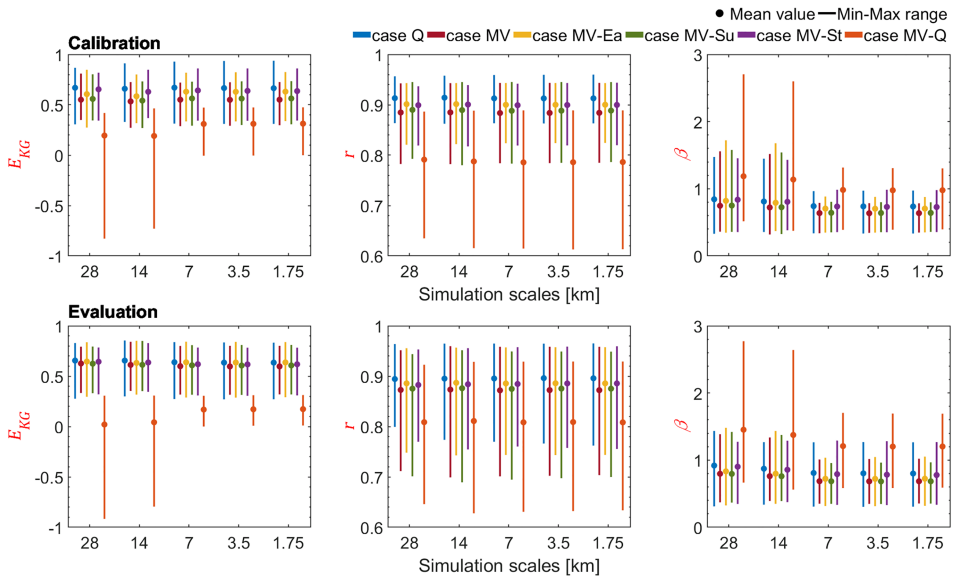

Figure 8. Statistics for model parameter transferability across spatial scales for streamflow performance for all calibration cases. The dots give the mean score, and the bars represent the min-max range of all values for 11 streamflow gauges. The colors correspond to the model calibration cases.

only Q compared to those in the semiarid northern part. These findings are in agreement with the study of Rakovec et al. (2016), which revealed a more pronounced sensitivity in E_a estimation in humid catchments in Europe through a multivariate calibration setting ($Q + S_t$). Similar results are obtained by Bai et al. (2016) when testing different E_p formulas in China. Contrastingly, Bai, Liu, and Liu (2018) found that their multivariate calibration setting ($Q + S_t$) benefitted more to E_a simulation in dry catchments than in wet catchments in China.

5.5. Parameter Transferability Across Spatial Scales

5.5.1. Streamflow Evaluation Across Spatial Scales

The model performance of streamflow in terms of scale transferability of the global parameters is given in Figure 8. The differences in model performance among calibration cases are conserved across spatial scales, with a median coefficient of variation of 1.6% for E_{KG}, 0.1% for r, and 6.4% for β.

5.5.2. Spatial Pattern Evaluation Across Spatial Scales

Long-term monthly maps of S_u (Figure 9a) and E_a (Figure 9b) are plotted along with T_s maps at various spatial resolutions. Here only the coarsest and finest resolutions (i.e., 28 and 1.75 km) are shown, but the same figures with intermediate resolutions and the spatial pattern efficiency are given in the supporting information (Figures S34–S37 and S32–S33).

The patterns of S_u is consistent with the patterns of T_s because the expectation is that the higher the T_s, the lower the S_u and vice versa (Figure 9a). For semiarid regions, E_a largely depends on water availability (i.e., rainfall) and is dominant for open water storages. In the VRB, E_a depicts an opposite pattern to T_s, which is shown in Figure 9b. The reproduction of both S_u and E_a in the multivariate calibration cases and across spatial scales show more plausible patterns with T_s, which are well preserved across scales with higher consistency, than their representation with case Q. The maps of the temporal correlation per grid cell are provided in the supporting information (Figures S35 and S37).

5.6. Benefit of Spatial Patterns and Data Types in Multivariate Calibration

5.6.1. Analysis of the Lake Volta Region

Evidence of the benefit of multivariate calibration with SRS data is exemplified in Figure 10 by zooming-in on the Volta Lake region in the southern part of the VRB (cf. Figure 1). Notwithstanding that mHM does not have a lake module, it is nicely noticeable that the model represents the heterogeneity in spatial patterns with the multivariate calibration cases being better than the Q-only calibration case. As it should be

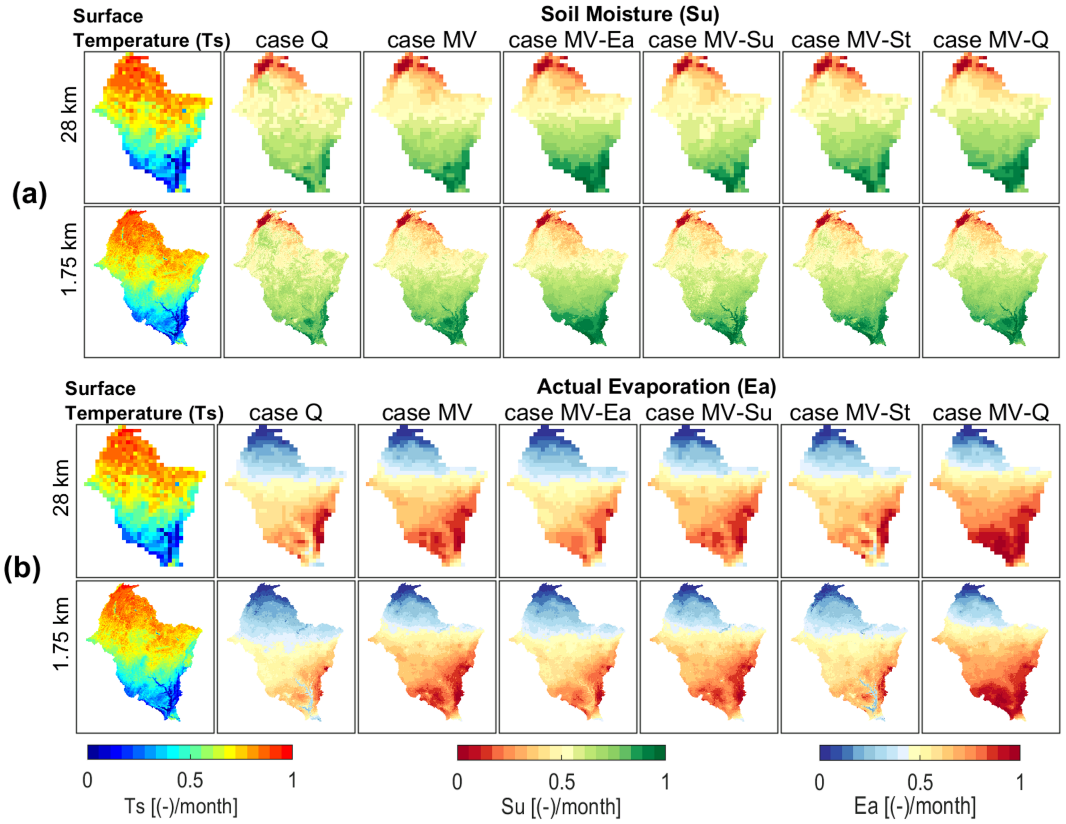

Figure 9. Long-term monthly average of land surface temperature compared to (a) soil moisture and (b) actual evaporation for all model calibration cases over the simulation period (2003–2012) at various spatial resolutions. The values are normalized for better emphasizing on patterns and using a unique color scale.

expected from the T_s patterns, case MV shows higher S_u and E_a over the Lake Volta but with lower E_a in its surroundings. This improvement in spatial patterns is not observed with case Q, confirming the limitations of the Q-only calibration and emphasizing the importance of patterns of SRS data for model calibration. Moreover, it can be inferred from the results that S_t is the most important variable for representing the lake area, while Q is the less critical variable.

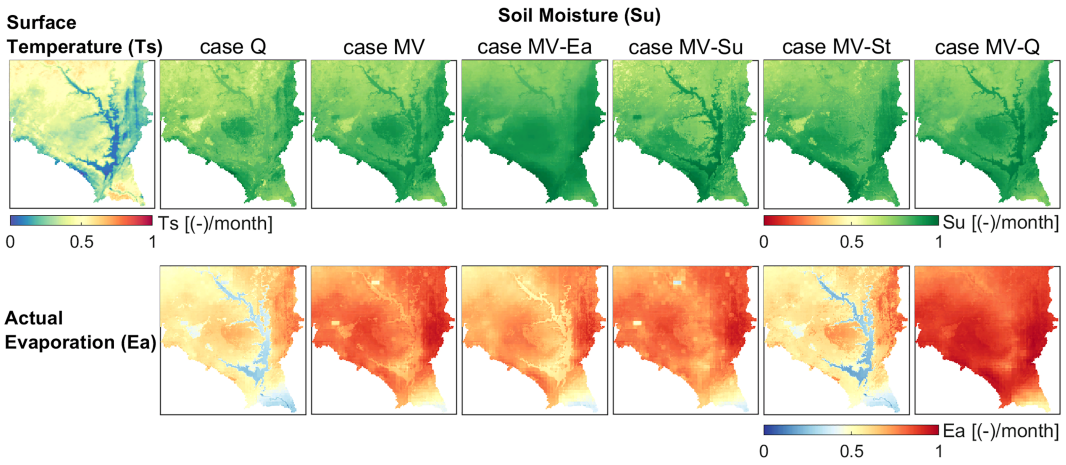

Figure 10. Comparison of the spatial patterns of (tow row) soil moisture and (bottom row) actual evaporation with land surface temperature (first map from left) in the Volta Lake region. The T_s map used as benchmark shows the Lake Volta depicted in dark blue with the lowest temperature in the region. The ability of the mHM model to highlight the lake area is assessed with the patterns of S_u and E_a for all model calibration cases.

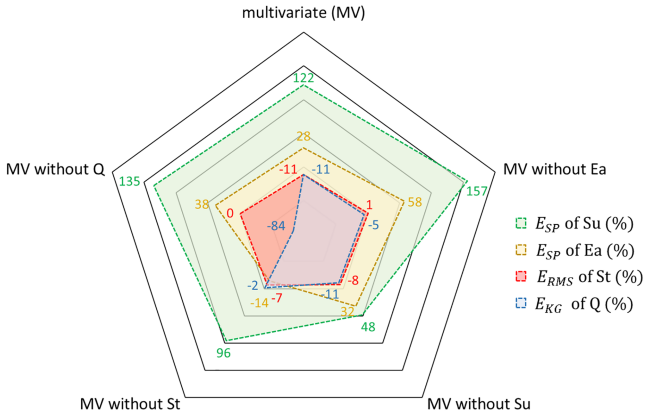

Figure 11. Relative difference in performance of multivariate (MV) calibration cases compared to Q-only calibration case. For every MV case, the values on the line from each vertex to the center of the polygon give the relative difference in performance with the Q-only calibration case, for all variables (i.e., Q, S_t, S_u, and E_a).

5.6.2. Comparison of the Multivariate Calibration Cases to the Benchmark

Figure 11 gives the gain or loss in model performance with different multivariate model calibration cases compared to the Q-only calibration case (i.e., the benchmark).

In general, the multivariate calibration cases show higher model predictive skill for many components of the hydrological system (i.e., Q, S_t, S_u, and E_a) when compared to the Q-only calibration. The decrease in the model performance for Q is usually counterbalanced with an increase in performance for S_t, S_u, and E_a, which might simply be the result of a solution to the artifacts caused by the Q-only calibration. These results reveal that the simulation of spatial patterns of S_u benefits most from the multivariate settings, followed by the simulation of E_a and S_t, while the decrease in Q performance varies widely depending on the multivariate calibration cases. A summary of the importance of each variable in predicting the others (i.e., Q, S_t, S_u, and E_a) is given in Table 4.

The most determinant variable for streamflow prediction in a multivariate setting is streamflow itself, similarly for S_u, while it is E_a for S_t, and S_t for E_a. Zeng and Cai (2016) also found that S_t controls the temporal variability of E_a. Surprisingly, E_a is the less critical variable for S_u prediction. However, it is worth stressing that only the spatial patterns of satellite E_a is exploited here. Moreover, E_a calculation in the model setup might be a reason of the limited contribution of satellite E_a in S_u prediction. The E_p calculation (cf. section 4.2) is done with time-variant and gridded leaf area index data that imposes heterogeneity on modeled E_a (Birhanu et al., 2019). Consequently, additional contribution from the satellite E_a in S_u prediction is expectedly limited in case the leaf area index data are in agreement with the satellite E_a. Moreover, not explicitly weighting the components of the multivariate objective function might have led to implicit weighting, which led to the artifact that some variables are not very good predictors for themselves.

5.7. Summary and Outlook

This work is a follow-up on several recent studies on multiobjective calibration and spatial pattern improvement in hydrological modeling (e.g., Demirel et al., 2018; Koch et al., 2018; Nijzink et al., 2018; Stisen et al., 2018; Yassin et al., 2017; Zink et al., 2018). The proposed multivariate calibration approach is a step forward in improving the realism of hydrological model predictions (Baroni et al., 2019; Clark et al., 2015; Rakovec et al., 2016) because not only a reliable temporal dynamic in the modeling objective but also plausible spatial patterns of several hydrological processes simultaneously are sought for. A key element of our study is the assessment of the plausibility of spatial patterns of soil moisture and evaporation with independent data of land surface temperature not used during the model calibration. With respect to the obtained performances, it can be concluded that spatial patterns of satellite data are a highly relevant and robust feature that can be used in multivariate calibration to improve the overall representation of the hydrological system even with trade-offs among the variables, which thereby confirms our research hypothesis.

A rigorous comparison of the proposed bias-insensitive metric with other spatial pattern metrics is left for future work. Further investigations can focus on setting a threshold for the acceptability of the modeled spatial patterns, which was not required here as the goal was to check the increase or decrease of spatial pattern performance rather than determining whether the patterns are good or bad in an absolute sense, when switching between streamflow-only and multivariate calibration cases.

Our methodology lacks in situ data for model evaluation, except streamflow. However, in situ measurements of soil moisture, evaporation, and terrestrial water storage at a large scale are rather rare (Vereecken et al.,

Table 4
Importance of Different Variables in Predicting Others in a Multivariate Calibration Setting

	Predictands					
	Temporal dynamics				Spatial patterns	
Predictors	Q	S_t	E_a	S_u	E_a	S_u
Q	++++	+	++	+++	+	+
S_t	++	++	++++	+++	++++	++
E_a	++	++++	+++	+	++	+
S_u	++	+	+	+++	+++	+++

Note. The degree of importance is as follows: low (+), moderate (++), high (+++), and very high (++++).

2008; Zink et al., 2018) and are also subject to uncertainties due to the nonuniformity of the data collection in space (Stisen et al., 2011). As we focus on spatial pattern assessment in this study, satellite data remain the only possible option for our large study area in West Africa, where ground measurements are a luxury (Dembélé et al., 2019).

The presented multivariate calibration reveals trade-offs among the objective functions for streamflow and for satellite data. However, trade-offs cannot be avoided as they originate from errors in input data, model structure, and lack of knowledge of the hydrological system (Bergström et al., 2002; Gupta et al., 1998; Yassin et al., 2017). Moreover, it was a deliberate choice to equally weight the components of the multivariate objective function (equation (4)) because no prior knowledge on the importance of each variable was available, and it was an objective of this study to know their contributions in the calibration procedure. In such situation, the default choice is to weight them equally (Bergström et al., 2002; Stisen et al., 2018). Weights are sometimes assigned to objective function components by iterative optimization testing different weights, which is, however, computationally demanding. It is also possible to transform the components of the multivariate objective functions to solve differences in their magnitudes (Madsen, 2003; Zink et al., 2018), but the effects of such transformations on the calibration procedure are unknown and they are not required if the metrics are dimensionless or of the same order of magnitude (Bergström et al., 2002). For completeness, Pareto plots showing the absolute values and the trade-off among the used objective functions are provided in the supporting information (Figures S38 and S39).

The climatic inputs influence somehow the spatial variability of the hydrological processes due to the aridity gradient in the VRB. The detailed results are valid for the VRB, but they can be generalized to regions with similar hydroclimatic characteristics. However, the applicability of the proposed multivariate calibration framework is, in principle, universal, as long as a DHM is used and spatial data sets are available. Further research can explore the applicability of the presented multivariate calibration strategies in different hydroclimatic regions with different spatial data sources, and different DHMs to understand how the model structure interacts with the performance of different calibration strategies. Choosing an adequate hydrological model (Addor & Melsen, 2019) is key to any good experiment. The MPR scheme used in mHM might have facilitated to some extent the reproduction of spatial patterns, but the MPR scheme can be similarly implemented with other models as demonstrated by previous studies (e.g., VIC and PCR-GLOBWB models; Mizukami et al., 2017; Samaniego et al., 2017). A sensitivity analysis to identify the model parameters that influence the representation of spatial patterns is a recommended outlook.

Future methodological developments could in particular focus on improved formulation of the multiobjective functions inspired by previous findings on the following topics: fitting of low flows and system signatures (Fowler, Peel, et al., 2018; Hrachowitz et al., 2014; Krause et al., 2005; Pushpalatha et al., 2012), gauge measurement weighting (Madsen, 2003), or subperiod calibration (Gharari et al., 2013). Additional key questions to address in this context include the model structural deficiencies (Gupta et al., 1998; Gupta et al., 2012) and the uncertainties of modeling data sets (i.e. input, calibration, and evaluation data), which can lead to erroneous model rejection (Beven, 2010, 2018, 2019b).

The above efforts in model improvement are particularly important for prediction in a changing environment (Fowler et al., 2018), and they can set avenues for prediction in ungauged basins solely from space.

6. Conclusion

This study presents a calibration approach using multiple data sources simultaneously, with the specificity of integrating only spatial patterns of satellite remote sensing data in the parameter estimation procedure. A bias-insensitive and multicomponent metric is proposed for spatial pattern matching. The study is carried out in the Volta River basin in West Africa. Results reveal the benefit of the multivariate calibration setting over the traditional calibration using only streamflow data. The main findings are as follows:

- Streamflow is a necessary variable, but alone it is not sufficient for reliably reproducing other hydrological fluxes and states.
- Spatial patterns of satellite data, without the absolute values, can be incorporated in the calibration procedure with bias-insensitive metrics.

- Multivariate calibration based on streamflow and satellite data can improve the overall representation of the hydrological system and thereby increase the model predictive skill.
- The reduction in streamflow performance in a multivariate setting is largely compensated by the gain in performance for other hydrological processes (i.e., terrestrial water storage, soil moisture, and evaporation).

We advocate for the adoption of multivariate calibration procedure focusing on spatial patterns in distributed hydrological models because it is a robust approach for addressing equifinality, reducing uncertainties, and enhancing the predictive skill of hydrological models in a changing environment.

Appendix A: Equations

Euclidean Distance

The Euclidian distance (D_E) between two points X and Y of coordinates ($x_1, x_2, ..., x_n$) and ($y_1, y_2, ..., y_n$) in an n-dimensional space (Upton & Cook, 2014) is given by

$$D_E = \sqrt{\sum_i^n (x_i - y_i)^2}. \tag{A1}$$

Nash-Sutcliffe Efficiencies

The Nash-Sutcliffe efficiency (Nash & Sutcliffe, 1970) of streamflow (E_{NS}) and the Nash-Sutcliffe efficiency of the logarithm of streamflow (E_{NSlog}) are formulated as follows:

$$E_{NS} = 1 - \frac{\sum_1^t (Q_{mod}(t) - Q_{obs}(t))^2}{\sum_1^t \left(Q_{obs}(t) - \overline{Q_{obs}}\right)^2} \tag{A2}$$

and

$$E_{NSlog} = 1 - \frac{\sum_1^t [\log(Q_{mod}(t)) - \log(Q_{obs}(t))]^2}{\sum_1^t \left[\log(Q_{obs}(t)) - \overline{\log(Q_{obs})}\right]^2}, \tag{A3}$$

where Q_{mod} and Q_{obs} are the modeled and observed streamflow, respectively, and t is the number of time steps. E_{NS} and E_{NSlog} range from $-\infty$ to their perfect score, that is, 1. These two metrics are used as skill scores to identify and discard parameter sets that provide implausible representations of the system and to subsequently better predict both high and low flows (Krause et al., 2005; Oudin et al., 2006; Pushpalatha et al., 2012; Santos et al., 2018).

z-Scores and Root-Mean-square Error

The standardized values (z-scores, Z_X) of X_{mod} and X_{obs} and the root-mean-square error (E_{RMS}) of a modeled variable (X_{mod}) and an observed variable (X_{obs}) of n elements, are defined as follows:

$$Z_X = \frac{X - \mu}{\sigma} \tag{A4}$$

and

$$E_{RMS}(X_{mod}, X_{obs}) = \sqrt{\frac{1}{n}\sum_1^n (X_{mod} - X_{obs})^2}. \tag{A5}$$

where μ and σ are the mean and the standard deviation of a given variable X, respectively.

Kling-Gupta efficiency

The Kling-Gupta efficiency (E_{KG}) was introduced by Gupta et al. (2009) and modified by Kling et al. (2012) to avoid some limitations of E_{NS}. E_{KG} combines correlation, bias, and variability measures and is defined as follows:

$$E_{KG} = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2}, \quad \text{(A6)}$$

where r is the Pearson correlation coefficient, β represents the bias term (i.e., ratio of means), and γ is the variability term (i.e., ratio of coefficients of variation).

Acknowledgments
The authors thank the providers of the data sets (see Table 2) used in this study. They thank the mHM team at CHS/UFZ (Germany), particularly Oldrich Rakovec and Pallav Kumar Shrestha, for guidance in setting up the model. They are grateful to the streamflow data providers and the National Hydrologic Services in the VBA member countries. The streamflow data sets were obtained from the Direction Générale des Ressources en Eau (DGRE) of Burkina Faso, the Hydrological Services Department (HSD) of Ghana, the Volta Basin Authority (VBA), and the Direction Générale de l'Eau et de l'Assainissement (DGEA) of Togo through a contact at the Université d'Abomey Calavi (UAC). Data available from the abovementioned data providers can be obtained from them. The streamflow data sets can be accessed from the Global Runoff Data Centre (www.bafg.de/GRDC). The model outputs are available at the Zenodo website (https://doi.org/10.5281/zenodo.3531873). The first author thanks the Swiss Confederation for financial support through the Swiss Government Excellence Scholarship (2016.0533/Burkina Faso/OP) and the Swiss National Science Foundation for the Doc.Mobility fellowship (SNF, P1LAP2_178071), under which this work was partly carried out at TU Delft (Netherlands). The last author is supported by a research grant from the Swiss National Science Foundation (SNF, PP00P2_157611). The authors thank Luis Samaniego (Associate Editor), Juraj Parajka, Simon Stisen, and the anonymous reviewer for their constructive comments that contributed in improving the final manuscript.

References

Addor, N., & Melsen, L. (2019). Legacy, rather than adequacy, drives the selection of hydrological models. *Water Resources Research*, *55*(1), 378–390. https://doi.org/10.1029/2018WR022958

AghaKouchak, A., Farahmand, A., Melton, F. S., Teixeira, J., Anderson, M. C., Wardlow, B. D., & Hain, C. R. (2015). Remote sensing of drought: Progress, challenges and opportunities. *Reviews of Geophysics*, *53*(2), 452–480. https://doi.org/10.1002/2014rg000456

Allen, R. G., L. S. Pereira, D. Raes, & M. Smith (1998). Crop evapotranspiration—Guidelines for computing crop water requirements—FAO Irrigation and drainage paper 56, 326 pp. http://academic.uprm.edu/abe/backup2/tomas/fao%2056.pdf

Andreini, M., N. Giesen, A. Edig, M. Fosu, and W. Andah (2000), Volta Basin water balance, 37 pp, University of Bonn, Center for Development Research (ZEF), https://ideas.repec.org/p/ags/ubzefd/280265.html

Bai, P., Liu, X., Zhang, Y., & Liu, C. (2018). Incorporating vegetation dynamics noticeably improved performance of hydrological model under vegetation greening. *Sci Total Environ*, *643*, 610–622. https://doi.org/10.1016/j.scitotenv.2018.06.233

Bai, P., Liu, X. M., & Liu, C. M. (2018). Improving hydrological simulations by incorporating GRACE data for model calibration. *Journal of Hydrology*, *557*, 291–304. https://doi.org/10.1016/j.jhydrol.2017.12.025

Bai, P., Liu, X. M., Yang, T. T., Li, F. D., Liang, K., Hu, S. S., & Liu, C. M. (2016). Assessment of the influences of different potential evapotranspiration inputs on the performance of monthly hydrological models under different climatic conditions. *Journal of Hydrometeorology*, *17*(8), 2259–2274. https://doi.org/10.1175/Jhm-D-15-0202.1

Balsamo, G., Agusti-Panareda, A., Albergel, C., Arduini, G., Beljaars, A., Bidlot, J., et al. (2018). Satellite and in situ observations for advancing global earth surface modelling: A review. *Remote Sensing*, *10*(12), 2038. https://doi.org/10.3390/rs10122038

Baroni, G., Schalge, B., Rakovec, O., Kumar, R., Schüler, L., Samaniego, L., et al. (2019). A comprehensive distributed hydrological modelling inter-comparison to support processes representation and data collection strategies. *Water Resources Research*, *55*(2), 990–1010. https://doi.org/10.1029/2018WR023941

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G. P., et al. (2017). Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrology and Earth System Sciences*, *21*(12), 6201–6217. https://doi.org/10.5194/hess-21-6201-2017

Becker, R., Koppa, A., Schulz, S., Usman, M., aus der Beek, T., & Schüth, C. (2019). Spatially distributed model calibration of a highly managed hydrological system using remote sensing-derived ET data. *Journal of Hydrology*, *577*, 123944. https://doi.org/10.1016/j.jhydrol.2019.123944

Bergström, S., Lindström, G., & Pettersson, A. (2002). Multi-variable parameter estimation to increase confidence in hydrological modelling. *Hydrological processes*, *16*(2), 413–421. https://doi.org/10.1002/hyp.332

Beven, K. (2006a). A manifesto for the equifinality thesis. *Journal of Hydrology*, *320*(1-2), 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007

Beven, K. (2006b). Searching for the Holy Grail of scientific hydrology: $Q_t = (S, R, \Delta t)A$ as closure. *Hydrology Earth System Sciences*, *10*(5), 609–618. https://doi.org/10.5194/hess-10-609-2006

Beven, K. (2010). Preferential flows and travel time distributions: Defining adequate hypothesis tests for hydrological process models. *Hydrological Processes*, *24*(12), 1537–1547. https://doi.org/10.1002/hyp.7718

Beven, K. (2018). On hypothesis testing in hydrology: Why falsification of models is still a really good idea. *Wiley Interdisciplinary Reviews-Water*, *5*(3), e1278. https://doi.org/10.1002/wat2.1278

Beven, K. (2019a). How to make advances in hydrological modelling. *Hydrology Research*, *50*(6), 1481–1494. https://doi.org/10.2166/nh.2019.134

Beven, K. (2019b). Towards a methodology for testing models as hypotheses in the inexact sciences. *Proceedings of the Royal Society A*, *475*(2224), 20180862). https://doi.org/10.1098/rspa.2018.0862

Beven, K., & Feyen, J. (2002). The future of distributed modelling—Special issue. *Hydrological Processes*, *16*(2), 169–172. https://doi.org/10.1002/hyp.325

Biondi, D., Freni, G., Iacobellis, V., Mascaro, G., & Montanari, A. (2012). Validation of hydrological models: Conceptual basis, methodological approaches and a proposal for a code of practice. *Physics and Chemistry of the Earth*, *42-44*, 70–76. https://doi.org/10.1016/j.pce.2011.07.037

Birhanu, D., Kim, H., & Jang, C. (2019). Effectiveness of introducing crop coefficient and leaf area index to enhance evapotranspiration simulations in hydrologic models. *Hydrological Processes*, hyp.13464. https://doi.org/10.1002/hyp.13464

Blöschl, G., Bierkens, M. F., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., et al. (2019). Twenty-three unsolved problems in hydrology (UPH)—A community perspective. *Hydrological Sciences Journal*, *64*(10), 1141–1158. https://doi.org/10.1080/02626667.2019.1620507

Blöschl, G., Sivapalan, M., Savenije, H., Wagener, T., & Viglione, A. (2013). *Runoff prediction in ungauged basins: Synthesis across processes, places and scales*, (p. 490). University Printing House, Cambridge CB2 8BS, United Kingdom: Cambridge University Press.

Boni, G., Entekhabi, D., & Castelli, F. (2001). Land data assimilation with satellite measurements for the estimation of surface energy balance components and surface control on evaporation. *Water resources research*, *37*(6), 1713–1722. https://doi.org/10.1029/2001WR900020

Bontemps, S., P. Defourny, E. V. Bogaert, O. Arino, V. Kalogirou, and J. R. Perez (2011), GLOBCOVER 2009—Products description and validation report, 53 pp. http://due.esrin.esa.int/files/GLOBCOVER2009_Validation_Report_2.2.pdf

Butler, D. (2014). Earth observation enters next phase. *Nature*, *508*(7495), 160–161. https://doi.org/10.1038/508160a

Cazenave, A., Champollion, N., Benveniste, J., & Chen, J. (2016). *Remote sensing and water resources* (p. 337). Springer International Publishing Switzerland: Springer. https://doi.org/10.1007/978-3-319-32449-4

Chen, L., & Wang, L. (2018). Recent advance in earth observation big data for hydrology. *Big Earth Data*, *2*(1), 86–107. https://doi.org/10.1080/20964471.2018.1435072

Chen, X., Long, D., Hong, Y., Zeng, C., & Yan, D. (2017). Improved modeling of snow and glacier melting by a progressive two-stage calibration strategy with GRACE and multisource data: How snow and glacier meltwater contributes to the runoff of the Upper Brahmaputra River basin? *Water Resources Research*, *53*(3), 2431–2466. https://doi.org/10.1002/2016WR019656

Clark, M. P., Bierkens, M. F. P., Samaniego, L., Woods, R. A., Uijlenhoet, R., Bennett, K. E., et al. (2017). The evolution of process-based hydrologic models: Historical challenges and the collective quest for physical realism. *Hydrology and Earth System Sciences*, *21*(7), 3427–3440. https://doi.org/10.5194/hess-21-3427-2017

Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., et al. (2015). Improving the representation of hydrologic processes in Earth system models. *Water Resources Research*, *51*(8), 5929–5956. https://doi.org/10.1002/2015wr017096

Clark, M. P., Schaefli, B., Schymanski, S. J., Samaniego, L., Luce, C. H., Jackson, B. M., et al. (2016). Improving the theoretical underpinnings of process-based hydrologic models. *Water Resources Research*, *52*(3), 2350–2365. https://doi.org/10.1002/2015WR017910

Cui, X., Guo, X., Wang, Y., Wang, X., Zhu, W., Shi, J., et al. (2019). Application of remote sensing to water environmental processes under a changing climate. *Journal of Hydrology*, *574*, 892–902. https://doi.org/10.1016/j.jhydrol.2019.04.078

Cui, Y., Chen, X., Gao, J., Yan, B., Tang, G., & Hong, Y. (2018). Global water cycle and remote sensing big data: Overview, challenge, and opportunities. *Big Earth Data*, *2*(3), 282–297. https://doi.org/10.1080/20964471.2018.1548052

Cunge, J. A. (1969). On the subject of a flood propagation computation method (Musklngum method). *Journal of Hydraulic Research*, *7*(2), 205–230. https://doi.org/10.1080/00221686909500264

Danielson, J. J., and D. B. Gesch (2011), Global multi-resolution terrain elevation data 2010 (GMTED2010) *Rep. 2331-1258*, 34 pp, US Geological Survey. https://doi.org/10.3133/ofr20111073

De Condappa, D., and J. Lemoalle (2009), Atlas de l'eau dans le basin de la Volta [Water atlas of the Volta Basin], 93 pp, http://hal.ird.fr/ird-00505116/document

Dembélé, M., Oriani, F., Tumbulto, J., Mariéthoz, G., & Schaefli, B. (2019). Gap-filling of daily streamflow time series using direct sampling in various hydroclimatic settings. *Journal of Hydrology*, *569*, 573–586. https://doi.org/10.1016/j.jhydrol.2018.11.076

Dembele, M., & Zwart, S. J. (2016). Evaluation and comparison of satellite-based rainfall products in Burkina Faso, West Africa. *International Journal of Remote Sensing*, *37*(17), 3995–4014. https://doi.org/10.1080/01431161.2016.1207258

Demirel, M. C., Mai, J., Mendiguren, G., Koch, J., Samaniego, L., & Stisen, S. (2018). Combining satellite data and appropriate objective functions for improved spatial pattern performance of a distributed hydrologic model. *Hydrology and Earth System Sciences*, *22*(2), 1299–1315. https://doi.org/10.5194/hess-22-1299-2018

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., et al. (2017). ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sensing of Environment*, *203*, 185–215. https://doi.org/10.1016/j.rse.2017.07.001

Efstratiadis, A., & Koutsoyiannis, D. (2010). One decade of multi-objective calibration approaches in hydrological modelling: A review. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, *55*(1), 58–78. https://doi.org/10.1080/02626660903526292

Ehlers, L. B., Sonnenborg, T. O., & Refsgaard, J. C. (2018). Observational and predictive uncertainties for multiple variables in a spatially distributed hydrological model. *Hydrological Processes*, hyp.13367. https://doi.org/10.1002/hyp.13367

Engman, E. T., & Gurney, R. J. (1991). *Remote sensing in hydrology*, (p. 225). Springer Netherlands: Chapman and Hall Ltd. https://www.springer.com/gp/book/9789401066709

Euser, T., Winsemius, H., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., & Savenije, H. (2013). A framework to assess the realism of model structures using hydrological signatures. *Hydrology Earth System Sciences*, *17*(5), 1893–1912. https://doi.org/10.5194/hess-17-1893-2013

Fatichi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., et al. (2016). An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *Journal of Hydrology*, *537*, 45–60. https://doi.org/10.1016/j.jhydrol.2016.03.026

Feddes, R. A., Kowalik, P., Kolinskamalinka, K., & Zaradny, H. (1976). Simulation of field water-uptake by plants using a soil-water dependent root extraction function. *Journal of Hydrology*, *31*(1-2), 13–26. https://doi.org/10.1016/0022-1694(76)90017-2

Fenicia, F., McDonnell, J. J., & Savenije, H. H. (2008). Learning from model improvement: On the contribution of complementary data to process understanding. *Water Resources Research*, *44*(6). https://doi.org/10.1029/2007WR006386

Food and Agriculture Organization/Global Information and Early Warning System (1998), Sahel weather and crop situation 1998, 8 pp, Food and Agriculture Organization (FAO)/Global Information and Early Warning System (GIEWS), http://www.fao.org/docrep/004/x0059e/x0059e00.htm.

Fovet, O., Ruiz, L., Hrachowitz, M., Faucheux, M., & Gascuel-Odoux, C. (2015). Hydrological hysteresis and its value for assessing process consistency in catchment conceptual models. *Hydrology Earth System Sciences*, *19*(1), 105–123. https://doi.org/10.5194/hess-19-105-2015

Fowler, K., Coxon, G., Freer, J., Peel, M., Wagener, T., Western, A., et al. (2018). Simulating runoff under changing climatic conditions: A framework for model improvement. *Water Resources Research*, *54*(12), 9812–9832. https://doi.org/10.1029/2018wr023989

Fowler, K., Peel, M., Western, A., & Zhang, L. (2018). Improved rainfall-runoff calibration for drying climate: Choice of objective function. *Water Resources Research*, *54*(5), 3392–3408. https://doi.org/10.1029/2017WR022466

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., et al. (2015). The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Scientific data*, *2*, 150066(1). https://doi.org/10.15780/G2RP4Q

Gash, J. (1979). An analytical model of rainfall interception by forests. *Quarterly Journal of the Royal Meteorological Society*, *105*(443), 43–55. https://doi.org/10.1002/qj.49710544304

Gharari, S., Hrachowitz, M., Fenicia, F., & Savenije, H. H. G. (2013). An approach to identify time consistent model parameters: Sub-period calibration. *Hydrology and Earth System Sciences*, *17*(1), 149–161. https://doi.org/10.5194/hess-17-149-2013

Gravetter, F. J., & Wallnau, L. B. (2013). Chapter 5, z-Scores: Location of scores and standardized distributions. In *Statistics for the behavioral sciences*, (pp. 137–162). Wadsworth, USA: Cengage Learning.

Grayson, R., & Bloschl, G. (2001). *Spatial patterns in catchment hydrology: Observations and modelling*, (p. 416). University Printing House, Cambridge CB2 8BS, United Kingdom: Cambridge University Press.

Gruber, A., Dorigo, W. A., Crow, W., & Wagner, W. (2017). Triple collocation-based merging of satellite soil moisture retrievals. *Ieee Transactions on Geoscience and Remote Sensing*, *55*(12), 6780–6792. https://doi.org/10.1109/Tgrs.2017.2734070

Guntner, A. (2008). Improvement of global hydrological models using GRACE data. *Surveys in Geophysics*, *29*(4-5), 375–397. https://doi.org/10.1007/s10712-008-9038-y

Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G., & Ye, M. (2012). Towards a comprehensive assessment of model structural adequacy. *Water Resources Research*, *48*(8). https://doi.org/10.1029/2011WR011044

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1-2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, *34*(4), 751–763. https://doi.org/10.1029/97wr03495

Gupta, H. V., Wagener, T., & Liu, Y. (2008). Reconciling theory with observations: Elements of a diagnostic approach to model evaluation. *Hydrological Processes: An International Journal*, *22*(18), 3802–3813. https://doi.org/10.1002/hyp.6989

Hapuarachchi, H. A. P., Wang, Q. J., & Pagano, T. C. (2011). A review of advances in flash flood forecasting. *Hydrological Processes*, *25*(18), 2771–2784. https://doi.org/10.1002/hyp.8040

Hargreaves, G. H., & Samani, Z. A. (1985). Reference crop evapotranspiration from temperature. *Applied Engineering in Agriculture*, *1*(2), 96–99. https://doi.org/10.13031/2013.26773

Hartmann, J., & Moosdorf, N. (2012). The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry Geophysics Geosystems*, *13*(12). https://doi.org/10.1029/2012gc004370

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One*, *12*(2), e0169748. https://doi.org/10.1371/journal.pone.0169748

Hrachowitz, M., & Clark, M. P. (2017). HESS opinions: The complementary merits of competing modelling philosophies in hydrology. *Hydrology Earth System Sciences*, *21*(8), 3953–3973. https://doi.org/10.5194/hess-21-3953-2017

Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., et al. (2014). Process consistency in models: The importance of system signatures, expert knowledge, and process complexity. *Water resources research*, *50*(9), 7445–7469. https://doi.org/10.1002/2014WR015484

Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J., Sivapalan, M., Pomeroy, J., et al. (2013). A decade of predictions in ungauged basins (PUB)—A review. *Hydrological Sciences Journal*, *58*(6), 1198–1255. https://doi.org/10.1080/02626667.2013.803183

Immerzeel, W. W., & Droogers, P. (2008). Calibration of a distributed hydrological model based on satellite evapotranspiration. *Journal of Hydrology*, *349*(3-4), 411–424. https://doi.org/10.1016/j.jhydrol.2007.11.017

Jiao, Y., Lei, H. M., Yang, D. W., Huang, M. Y., Liu, D. F., & Yuan, X. (2017). Impact of vegetation dynamics on hydrological processes in a semi-arid basin by using a land surface-hydrology coupled model. *Journal of Hydrology*, *551*, 116–131. https://doi.org/10.1016/j.jhydrol.2017.05.060

Kampf, S. K., & Burges, S. J. (2007). A framework for classifying and comparing distributed hillslope and catchment hydrologic models. *Water resources research*, *43*(5). https://doi.org/10.1029/2006WR005370

Khu, S. T., & Madsen, H. (2005). Multiobjective calibration with Pareto preference ordering: An application to rainfall-runoff model calibration. *Water Resources Research*, *41*(3). https://doi.org/10.1029/2004WR003041

Klemas, V. (2014). Remote sensing of floods and flood-prone areas: An overview. *Journal of Coastal Research*, *314*(4), 1005–1013. https://doi.org/10.2112/JCOASTRES-D-14-00160.1

Klemes, V. (1986). Operational testing of hydrological simulation-models. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, *31*(1), 13–24. https://doi.org/10.1080/02626668609491024

Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, *424-425*, 264–277. https://doi.org/10.1016/j.jhydrol.2012.01.011

Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. *Hydrology and Earth System Sciences Discussions*, *1-7*. https://doi.org/10.5194/hess-2019-327

Knoche, M., Fischer, C., Pohl, E., Krause, P., & Merz, R. (2014). Combined uncertainty of hydrological model complexity and satellite-based forcing data evaluated in two data-scarce semi-arid catchments in Ethiopia. *Journal of Hydrology*, *519*, 2049–2066. https://doi.org/10.1016/j.jhydrol.2014.10.003

Ko, A., Mascaro, G., & Vivoni, E. R. (2019). Strategies to improve and evaluate physics-based hyperresolution hydrologic simulations at regional basin scales. *Water Resources Research*, *55*(2), 1129–1152. https://doi.org/10.1029/2018WR023521

Koch, J., Demirel, M. C., & Stisen, S. (2018). The SPAtial EFficiency metric (SPAEF): Multiple-component evaluation of spatial patterns for optimization of hydrological models. *Geoscientific Model Development*, *11*(5), 1873–1886. https://doi.org/10.5194/gmd-11-1873-2018

Koch, J., Jensen, K. H., & Stisen, S. (2015). Toward a true spatial model evaluation in distributed hydrological modeling: Kappa statistics, fuzzy theory, and EOF-analysis benchmarked by the human perception and evaluated against a modeling case study. *Water Resources Research*, *51*(2), 1225–1246. https://doi.org/10.1002/2014WR016607

Koppa, A., Gebremichael, M., & Yeh, W. W. (2019). Multivariate calibration of large scale hydrologic models: The necessity and value of a Pareto optimal approach. *Advances in Water Resources*, *130*, 129–146. https://doi.org/10.1016/j.advwatres.2019.06.005

Krause, P., Boyle, D., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in geosciences*, *5*, 89–97. https://doi.org/10.5194/adgeo-5-89-2005

Kumar, R., Samaniego, L., & Attinger, S. (2013). Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations. *Water Resources Research*, *49*(1), 360–379. https://doi.org/10.1029/2012wr012195

Lakshmi, V. (2000). A simple surface temperature assimilation scheme for use in land surface models. *Water Resources Research*, *36*(12), 3687–3700. https://doi.org/10.1029/2000WR900204

Lakshmi, V., Jackson, T. J., & Zehrfuhs, D. (2003). Soil moisture–temperature relationships: Results from two field experiments. *Hydrological processes*, *17*(15), 3041–3057. https://doi.org/10.1002/hyp.1275

Landerer, F. W., & Swenson, S. C. (2012). Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resources Research*, *48*(4). https://doi.org/10.1029/2011wr011453

Legates, D. R., & McCabe, G. J. (1999). Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, *35*(1), 233–241. https://doi.org/10.1029/1998wr900018

Leroux, D. J., Pellarin, T., Vischel, T., Cohard, J.-M., Gascon, T., Gibon, F., et al. (2016). Assimilation of SMOS soil moisture into a distributed hydrological model and impacts on the water cycle variables over the Ouémé catchment in Benin. *Hydrology and Earth System Sciences*, *20*(7), 2827–2840. https://doi.org/10.5194/hess-20-2827-2016

Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M., & Wood, E. F. (2015). Inroads of remote sensing into hydrologic science during the WRR era. *Water Resources Research*, *51*(9), 7309–7342. https://doi.org/10.1002/2015wr017616

Li, B. L., Rodell, M., Zaitchik, B. F., Reichle, R. H., Koster, R. D., & van Dam, T. M. (2012). Assimilation of GRACE terrestrial water storage into a land surface model: Evaluation and potential value for drought monitoring in western and central Europe. *Journal of Hydrology*, *446-447*, 103–115. https://doi.org/10.1016/j.jhydrol.2012.04.035

Li, Y., Grimaldi, S., Pauwels, V. R., & Walker, J. P. (2018). Hydrologic model calibration using remotely sensed soil moisture and discharge measurements: The impact on predictions at gauged and ungauged locations. *Journal of hydrology*, *557*, 897–909. https://doi.org/10.1016/j.jhydrol.2018.01.013

Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H. J., Kumar, S., Moradkhani, H., et al. (2012). Advancing data assimilation in operational hydrologic forecasting: Progresses, challenges, and emerging opportunities. *Hydrology and Earth System Sciences*, *16*(10), 3863–3887. https://doi.org/10.5194/hess-16-3863-2012

Liu, Y. Y., Dorigo, W. A., Parinussa, R. M., de Jeu, R. A. M., Wagner, W., McCabe, M. F., et al. (2012). Trend-preserving blending of passive and active microwave soil moisture retrievals. *Remote Sensing of Environment*, *123*, 280–297. https://doi.org/10.1016/j.rse.2012.03.014

Livneh, B., & Lettenmaier, D. P. (2012). Multi-criteria parameter estimation for the unified land model. *Hydrology and Earth System Sciences*, *16*(8), 3029–3048. https://doi.org/10.5194/hess-16-3029-2012

Lopez, P. L., Sutanudjaja, E. H., Schellekens, J., Sterk, G., & Bierkens, M. F. P. (2017). Calibration of a large-scale hydrological model using satellite-based soil moisture and evapotranspiration products. *Hydrology and Earth System Sciences*, *21*(6), 3125–3144. https://doi.org/10.5194/hess-21-3125-2017

Ma, Q. M., Xiong, L. H., Liu, D. D., Xu, C. Y., & Guo, S. L. (2018). Evaluating the temporal dynamics of uncertainty contribution from satellite precipitation input in rainfall-runoff modeling using the variance decomposition method. *Remote Sensing*, *10*(12), 1876. https://doi.org/10.3390/rs10121876

Madsen, H. (2003). Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. *Advances in Water Resources*, *26*(2), 205–216. https://doi.org/10.1016/S0309-1708(02)00092-1

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernandez-Prieto, D., et al. (2017). GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development*, *10*(5), 1903–1925. https://doi.org/10.5194/gmd-10-1903-2017

McCabe, M. F., Rodell, M., Alsdorf, D. E., Miralles, D. G., Uijlenhoet, R., Wagner, W., et al. (2017). The future of Earth observation in hydrology. *Hydrol Earth Syst Sci*, *21*(7), 3879–3914. https://doi.org/10.5194/hess-21-3879-2017

McDonnell, J., Sivapalan, M., Vaché, K., Dunn, S., Grant, G., Haggerty, R., et al. (2007). Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology. *Water Resources Research*, *43*(7). https://doi.org/10.1029/2006WR005467

Mendiguren, G., Koch, J., & Stisen, S. (2017). Spatial pattern evaluation of a calibrated national hydrological model—A remote-sensing-based diagnostic approach. *Hydrology and Earth System Sciences*, *21*(12), 5987–6005. https://doi.org/10.5194/hess-21-5987-2017

Mendoza, M., Bocco, G., & Bravo, M. (2002). Spatial prediction in hydrology: Status and implications in the estimation of hydrological processes for applied research. *Progress in Physical Geography*, *26*(3), 319–338. https://doi.org/10.1191/0309133302pp335ra

Minville, M., Cartier, D., Guay, C., Leclaire, L. A., Audet, C., Le Digabel, S., & Merleau, J. (2014). Improving process representation in conceptual hydrological model calibration using climate simulations. *Water Resources Research*, *50*(6), 5044–5073. https://doi.org/10.1002/2013wr013857

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., & Dolman, A. J. (2011). Global land-surface evaporation estimated from satellite-based observations. *Hydrology and Earth System Sciences*, *15*(2), 453–469. https://doi.org/10.5194/hess-15-453-2011

Mizukami, N., Clark, M. P., Newman, A. J., Wood, A. W., Gutmann, E. D., Nijssen, B., et al. (2017). Towards seamless large-domain parameter estimation for hydrologic models. *Water Resources Research*, *53*(9), 8020–8040. https://doi.org/10.1002/2017wr020401

Montanari, A., Bahr, J., Bloschl, G., Cai, X. M., Mackay, D. S., Michalak, A. M., et al. (2015). Fifty years of *Water Resources Research*: Legacy and perspectives for the science of hydrology. *Water Resources Research*, *51*(9), 6797–6803. https://doi.org/10.1002/2015wr017998

Mul, M., E. Obuobie, R. Appoh, K. Kankam-Yeboah, E. Bekoe-Obeng, B. Amisigo, et al. (2015), Water resources assessment of the Volta River basin *Rep. 9290908297*, 82 pp, International Water Management Institute (IWMI), http://www.iwmi.cgiar.org/Publications/Working_Papers/working/wor166.pdf.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

National Academies of Sciences, Engineering, and Medicine (2019). *Thriving on our changing planet: A decadal strategy for Earth observation from space*, edited by, *E. National Academies of Sciences, Engineering, and Medicine (NASEM)*, (p. 716). Washington, DC: The National Academies Press. https://doi.org/10.17226/24938

National Aeronautics and Space Administration (2019), Time average removed from monthly solutions, GRACE monthly mass grids—Land, Jet Propulsion Laboratory, http://grace.jpl.nasa.gov/data/get-data/monthly-mass-grids-land/, [accessed 2019.07.20].

Nijzink, R. C., Almeida, S., Pechlivanidis, I. G., Capell, R., Gustafssons, D., Arheimer, B., et al. (2018). Constraining conceptual hydrological models with multiple information sources. *Water Resources Research*, *54*(10), 8332–8362. https://doi.org/10.1029/2017wr021895

Oudin, L., Andréassian, V., Mathevet, T., Perrin, C., & Michel, C. (2006). Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations. *Water Resources Research*, *42*(7). https://doi.org/10.1029/2005WR004636

Oyana, T. J., & Margai, F. (2015). Chapter 3, Using statistical measures to analyze data distributions. In *Spatial analysis: Statistics, visualization, and computational methods*, (pp. 55–86). CRC Press 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487, USA: CRC Press.

Paniconi, C., & Putti, M. (2015). Physically based modeling in catchment hydrology at 50: Survey and outlook. *Water Resources Research*, *51*(9), 7090–7129. https://doi.org/10.1002/2015wr017780

Parajka, J., Naeimi, V., Bloschl, G., & Komma, J. (2009). Matching ERS scatterometer based soil moisture patterns with simulations of a conceptual dual layer hydrologic model over Austria. *Hydrology and Earth System Sciences*, *13*(2), 259–271. https://doi.org/10.5194/hess-13-259-2009

Pasetto, D., Arenas-Castro, S., Bustamante, J., Casagrandi, R., Chrysoulakis, N., Cord, A. F., et al. (2018). Integration of satellite remote sensing data in ecosystem modelling at local scales: Practices and trends. *Methods in Ecology and Evolution*, *9*(8), 1810–1821. https://doi.org/10.1111/2041-210x.13018

Peters-Lidard, C. D., Clark, M., Samaniego, L., Verhoest, N. E., Van Emmerik, T., Uijlenhoet, R., et al. (2017). Scaling, similarity, and the fourth paradigm for hydrology. *Hydrology earth system sciences*, *21*(7), 3701–3713. https://doi.org/10.5194/hess-21-3701-2017

Pokhrel, P., Gupta, H. V., & Wagener, T. (2008). A spatial regularization approach to parameter estimation for a distributed watershed model. *Water Resources Research*, *44*(12). https://doi.org/10.1029/2007wr006615

Pomeon, T., Diekkruger, B., & Kumar, R. (2018). Computationally efficient multivariate calibration and validation of a grid-based hydrologic model in sparsely gauged West African river basins. *Water*, *10*(10), 1418. https://doi.org/10.3390/w10101418

Pool, S., Vis, M., & Seibert, J. (2018). Evaluating model performance: Towards a non-parametric variant of the Kling-Gupta efficiency. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, *63*(13-14), 1941–1953. https://doi.org/10.1080/02626667.2018.1552002

Priestley, C. H. B., & Taylor, R. (1972). On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly weather review*, *100*(2), 81–92. https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2

Pushpalatha, R., Perrin, C., Le Moine, N., & Andreassian, V. (2012). A review of efficiency criteria suitable for evaluating low-flow simulations. *Journal of Hydrology*, *420-421*, 171–182. https://doi.org/10.1016/j.jhydrol.2011.11.055

Rajib, A., Evenson, G. R., Golden, H. E., & Lane, C. R. (2018). Hydrologic model predictability improves with spatially explicit calibration using remotely sensed evapotranspiration and biophysical parameters. *Journal of hydrology*, *567*, 668–683. https://doi.org/10.1016/j.jhydrol.2018.10.024

Rajib, A., Merwade, V., & Yu, Z. (2018). Rationale and efficacy of assimilating remotely sensed potential evapotranspiration for reduced uncertainty of hydrologic models. *Water Resources Research*, *54*(7), 4615–4637. https://doi.org/10.1029/2017WR021147

Rakovec, O., Kumar, R., Attinger, S., & Samaniego, L. (2016). Improving the realism of hydrologic model functioning through multivariate parameter estimation. *Water Resources Research*, *52*(10), 7779–7792. https://doi.org/10.1002/2016wr019430

Refsgaard, J. C. (2001). Towards a formal approach to calibration and validation of models using spatial data. In *Spatial patterns in catchment hydrology: observations and modelling* (pp. 329–354). University Printing House, Cambridge CB2 8BS, United Kingdom: Cambridge University Press.

Revilla-Romero, B., Hirpa, F. A., Thielen-del Pozo, J., Salamon, P., Brakenridge, R., Pappenberger, F., & De Groeve, T. (2015). On the use of global flood forecasts and satellite-derived inundation maps for flood monitoring in data-sparse regions. *Remote Sensing*, *7*(11), 15,702–15,728. https://doi.org/10.3390/rs71115702

Rientjes, T., Muthuwatta, L. P., Bos, M., Booij, M. J., & Bhatti, H. (2013). Multi-variable calibration of a semi-distributed hydrological model using streamflow data and satellite-based evapotranspiration. *Journal of hydrology*, *505*, 276–290. https://doi.org/10.1016/j.jhydrol.2013.10.006

Sakumura, C., Bettadpur, S., & Bruinsma, S. (2014). Ensemble prediction and intercomparison analysis of GRACE time-variable gravity field models. *Geophysical Research Letters*, *41*(5), 1389–1397. https://doi.org/10.1002/2013GL058632

Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, *46*(5). https://doi.org/10.1029/2008WR007327

Samaniego, L., Kumar, R., Thober, S., Rakovec, O., Zink, M., Wanders, N., et al. (2017). Toward seamless hydrologic predictions across spatial scales. *Hydrology and Earth System Sciences*, *21*(9), 4323–4346. https://doi.org/10.5194/hess-21-4323-2017

Santos, L., Thirel, G., & Perrin, C. (2018). Technical note: Pitfalls in using log-transformed flows within the KGE criterion. *Hydrology and Earth System Sciences*, *22*(8), 4583–4591. https://doi.org/10.5194/hess-22-4583-2018

Savenije, H. H. (2001). Equifinality, a blessing in disguise? *Hydrological Processes*, *15*(14), 2835–2838. https://doi.org/10.1002/hyp.494

Schaefli, B., & Gupta, H. V. (2007). Do Nash values have value? *Hydrological Processes*, *21*(15), 2075–2080. https://doi.org/10.1002/hyp.6825

Schmugge, T. J., Kustas, W. P., Ritchie, J. C., Jackson, T. J., & Rango, A. (2002). Remote sensing in hydrology. *Advances in Water Resources*, *25*(8-12), 1367–1385. https://doi.org/10.1016/S0309-1708(02)00065-9

Schoups, G., Hopmans, J. W., Young, C., Vrugt, J., & Wallender, W. W. (2005). Multi-criteria optimization of a regional spatially-distributed subsurface water flow model. *Journal of Hydrology*, *311*(1-4), 20–48. https://doi.org/10.1016/j.jhydrol.2005.01.001

Schultz, G. A., & Engman, E. T. (2012). *Remote sensing in hydrology and water management*, (p. 483). Springer-Verlag Berlin Heidelberg: Springer Science & Business Media.

Semenova, O., & Beven, K. (2015). Barriers to progress in distributed hydrological modelling. *Hydrological Processes*, *29*(8), 2074–2078. https://doi.org/10.1002/hyp.10434

Senay, G. B., Velpuri, N. M., Bohms, S., Budde, M., Young, C., Rowland, J., & Verdin, J. (2015). Drought monitoring and assessment: remote sensing and modeling approaches for the famine early warning systems network. In *Hydro-Meteorological Hazards, Risks and Disasters* (pp. 233–262). The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK: Elsevier. https://doi.org/10.1016/C2011-0-07025-4

Shafii, M., & Tolson, B. A. (2015). Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. *Water Resources Research*, *51*(5), 3796–3814. https://doi.org/10.1002/2014wr016520

Sheffield, J., Wood, E., Pan, M., Beck, H., Coccia, G., Serrat-Capdevila, A., & Verbist, K. (2018). Satellite remote sensing for water resources management: Potential for supporting sustainable development in data-poor regions. *Water Resources Research*, *54*(12), 9724–9758. https://doi.org/10.1029/2017WR022437

Sheffield, J., Wood, E. F., & Roderick, M. L. (2012). Little change in global drought over the past 60 years. *Nature*, *491*(7424), 435–438. https://doi.org/10.1038/nature11575

Silvestro, F., Gabellani, S., Rudari, R., Delogu, F., Laiolo, P., & Boni, G. (2015). Uncertainty reduction and parameter estimation of a distributed hydrological model with ground and remote-sensing data. *Hydrology and Earth System Sciences*, *19*(4), 1727–1751. https://doi.org/10.5194/hess-19-1727-2015

Singh, V. P. (2018). Hydrologic modeling: Progress and future directions. *Geoscience Letters*, *5*(1), 1–18. https://doi.org/10.1186/s40562-018-0113-z

Sivapalan, M. (2003). Prediction in ungauged basins: A grand challenge for theoretical hydrology. *Hydrological Processes*, *17*(15), 3163–3170. https://doi.org/10.1002/hyp.5155

Spaaks, J. H., & Bouten, W. (2013). Resolving structural errors in a spatially distributed hydrologic model using ensemble Kalman filter state updates. *Hydrology and Earth System Sciences*, *17*(9), 3455–3472. https://doi.org/10.5194/hess-17-3455-2013

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, *15*, 72–101. https://doi.org/10.2307/1412159

Stisen, S., Koch, J., Sonnenborg, T. O., Refsgaard, J. C., Bircher, S., Ringgaard, R., & Jensen, K. H. (2018). Moving beyond run-off calibration—Multivariable optimization of a surface–subsurface–atmosphere model. *Hydrological Processes*, *32*(17), 2654–2668. https://doi.org/10.1002/hyp.13177

Stisen, S., McCabe, M. F., Refsgaard, J. C., Lerer, S., & Butts, M. B. (2011). Model parameter analysis using remotely sensed pattern information in a multi-constraint framework. *Journal of Hydrology*, *409*(1-2), 337–349. https://doi.org/10.1016/j.jhydrol.2011.08.030

Su, Z., He, Y., Dong, X., & Wang, L. (2017). Drought monitoring and assessment using remote sensing. In *Remote Sensing of Hydrological Extremes* (pp. 151–172). Springer International Publishing Switzerland: Springer.

Sutanudjaja, E., Van Beek, L., De Jong, S., Van Geer, F., & Bierkens, M. (2014). Calibrating a large-extent high-resolution coupled groundwater-land surface model using soil moisture and discharge data. *Water Resources Research*, *50*(1), 687–705. https://doi.org/10.1002/2013WR013807

Swenson, S. C. (2012), GRACE monthly land water mass grids NETCDF release 5.0. Ver. 5.0. PO.DAAC, CA, USA, https://doi.org/10.5067/TELND-NC005, [Dataset accessed 2018.11.01].

Tang, Q. H., Gao, H. L., Lu, H., & Lettenmaier, D. P. (2009). Remote sensing: Hydrology. *Progress in Physical Geography*, *33*(4), 490–509. https://doi.org/10.1177/0309133309346650

Tangdamrongsub, N., Steele-Dunne, S. C., Gunter, B. C., Ditmar, P. G., Sutanudjaja, E. H., Sun, Y., et al. (2017). Improving estimates of water resources in a semi-arid region by assimilating GRACE data into the PCR-GLOBWB hydrological model. *Hydrology and Earth System Sciences*, *21*(4), 2053–2074. https://doi.org/10.5194/hess-21-2053-2017

Tapley, B. D., Bettadpur, S., Watkins, M., & Reigber, C. (2004). The gravity recovery and climate experiment: Mission overview and early results. *Geophysical Research Letters*, *31*(9). https://doi.org/10.1029/2004gl019920

Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F. W., Dutta, D., & Kim, S. (2017). Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. *Environmental Modelling & Software*, *90*, 201–216. https://doi.org/10.1016/j.envsoft.2017.01.006

Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., & De Roo, A. (2013). Hydrological evaluation of satellite-based rainfall estimates over the Volta and Baro-Akobo Basin. *Journal of Hydrology*, *499*, 324–338. https://doi.org/10.1016/j.jhydrol.2013.07.012

Thober, S., M. Cuntz, M. Kelbling, R. Kumar, J. Mai, and L. J. G. M. D. D. Samaniego (2019), The multiscale routing model mRM v1. 0: Simple river routing at resolutions from 1 to 50 km, *2019*, 1-26, https://doi.org/10.5194/gmd-12-2501-2019.

Tian, S. Y., Tregoning, P., Renzullo, L. J., van Dijk, A. I. J. M., Walker, J. P., Pauwels, V. R. N., & Allgeyer, S. (2017). Improved water balance component estimates through joint assimilation of GRACE water storage and SMOS soil moisture retrievals. *Water Resources Research*, *53*(3), 1820–1840. https://doi.org/10.1002/2016wr019641

Tobin, K. J., & Bennett, M. E. (2017). Constraining SWAT calibration with remotely sensed evapotranspiration data. *JAWRA Journal of the American Water Resources Association*, *53*(3), 593–604. https://doi.org/10.1111/1752-1688.12516

Tolson, B. A., & Shoemaker, C. A. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, *43*(1). https://doi.org/10.1029/2005wr004723

Trabucco, A., and R. Zomer (2018), Global aridity index and potential evapotranspiration (ET0) climate database v2. CGIAR Consortium for Spatial Information (CGIAR-CSI). Published online, available from the CGIAR-CSI GeoPortal at https://cgiarcsi.community accessed 11.07.2019].

Tucker, C. J., Pinzon, J. E., Brown, M. E., Slayback, D. A., Pak, E. W., Mahoney, R., et al. (2005). An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *International Journal of Remote Sensing*, *26*(20), 4485–4498. https://doi.org/10.1080/01431160500168686

United Nations Environment Programme (1997), World atlas of desertification/co-ordinating editors, Nick Middleton and David Thomas, 182 pp, United Nations Environment Programme, https://digitallibrary.un.org/record/245955.

Upton, G., & Cook, I. (2014). *A dictionary of statistics*. Oxford, United Kingdom: Oxford University Press.

Valente, F., David, J. S., & Gash, J. H. C. (1997). Modelling interception loss for two sparse eucalypt and pine forests in central Portugal using reformulated Rutter and Gash analytical models. *Journal of Hydrology*, *190*(1-2), 141–162. https://doi.org/10.1016/S0022-1694(96)03066-1

Vereecken, H., Huisman, J. A., Bogena, H., Vanderborght, J., Vrugt, J. A., & Hopmans, J. W. (2008). On the value of soil moisture measurements in vadose zone hydrology: A review. *Water Resources Research*, *44*(4). https://doi.org/10.1029/2008wr006829

Wagner, W., W. Dorigo, R. de Jeu, D. Fernandez, J. Benveniste, E. Haas, and M. Ertl (2012), Fusion of active and passive microwave observations to create an essential climate variable data record on soil moisture, paper presented at ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Annals).

Wambura, F. J., Dietrich, O., & Lischeid, G. (2018). Improving a distributed hydrological model using evapotranspiration-related boundary conditions as additional constraints in a data-scarce river basin. *Hydrological processes*, *32*(6), 759–775. https://doi.org/10.1002/hyp.11453

Wan, Z., S. Hook, and G. Hulley (2015), MYD11A2 MODIS/Aqua Land Surface Temperature/Emissivity 8-Day L3 Global 1km SIN Grid V00610.5067/MODIS/MYD11A2.006, [Dataset accessed 2019.02.13].

Wanders, N., Bierkens, M. F., de Jong, S. M., de Roo, A., & Karssenberg, D. (2014). The benefits of using remotely sensed soil moisture in parameter identification of large-scale hydrological models. *Water Resources Research*, *50*(8), 6874–6891. https://doi.org/10.1002/2013WR014639

Wang, L., Qu, J., Zhang, S., Hao, X., & Dasgupta, S. (2007). Soil moisture estimation using MODIS and ground measurements in eastern China. *International Journal of Remote Sensing*, *28*(6), 1413–1418. https://doi.org/10.1080/01431160601075525

Wealands, S. R., Grayson, R. B., & Walker, J. P. (2005). Quantitative comparison of spatial fields for hydrological model assessment––Some promising approaches. *Advances in Water Resources*, *28*(1), 15–32. https://doi.org/10.1016/j.advwatres.2004.10.001

Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., & Viterbo, P. (2014). The WFDEI meteorological forcing data set: WATCH Forcing data methodology applied to ERA-Interim reanalysis data. *Water Resources Research*, *50*(9), 7505–7514. https://doi.org/10.1002/2014WR015638

Werth, S., Guntner, A., Petrovic, S., & Schmidt, R. (2009). Integration of GRACE mass variations into a global hydrological model. *Earth and Planetary Science Letters*, *277*(1-2), 166–173. https://doi.org/10.1016/j.epsl.2008.10.021

Williams, T. O., Mul, M. L., Biney, C. A., & Smakhtin, V. (2016). *The Volta River basin: Water for food, economic growth and environment*, (p. 282). London, UK: Routledge. https://doi.org/10.4324/9781315707334

Wu, H., Adler, R. F., Tian, Y., Huffman, G. J., Li, H., & Wang, J. (2014). Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model. *Water Resources Research*, *50*(3), 2693–2717. https://doi.org/10.1002/2013WR014710

Wulder, M. A., & Coops, N. C. (2014). Satellites: Make Earth observations open access. *Nature News*, *513*(7516), 30. https://doi.org/10.1038/513030a

Xu, X. Y., Li, J., & Tolson, B. A. (2014). Progress in integrating remote sensing data and hydrologic modeling. *Progress in Physical Geography*, *38*(4), 464–498. https://doi.org/10.1177/0309133314536583

Yassin, F., Razavi, S., Wheater, H., Sapriza-Azuri, G., Davison, B., & Pietroniro, A. (2017). Enhanced identification of a hydrologic model using streamflow and satellite water storage data: A multicriteria sensitivity analysis and optimization approach. *Hydrological Processes*, *31*(19), 3320–3333. https://doi.org/10.1002/hyp.11267

Zeng, R. J., & Cai, X. M. (2016). Climatic and terrestrial storage control on evapotranspiration temporal variability: Analysis of river basins around the world. *Geophysical Research Letters*, *43*(1), 185–195. https://doi.org/10.1002/2015gl066470

Zhu, Z. C., Bi, J., Pan, Y. Z., Ganguly, S., Anav, A., Xu, L., et al. (2013). Global data sets of vegetation leaf area index (LAI)3g and fraction of photosynthetically active radiation (FPAR)3g derived from Global Inventory Modeling and Mapping Studies (GIMMS) normalized difference vegetation index (NDVI3g) for the period 1981 to 2011. *Remote Sensing*, *5*(2), 927–948. https://doi.org/10.3390/rs5020927

Zink, M., Mai, J., Cuntz, M., & Samaniego, L. (2018). Conditioning a hydrologic model using patterns of remotely sensed land surface temperature. *Water Resources Research*, *54*(4), 2976–2998. https://doi.org/10.1002/2017wr021346